ENCYCLOPEDIA OF ACOUSTICS

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker
Copyright © 1997 John Wiley & Sons, Inc.

EDITORIAL BOARD

Malcolm J. Crocker, *Editor-in-Chief*

Robert E. Apfel, Yale University, New Haven, CT
David T. Blackstock, University of Texas, Austin, TX
Leo L. Beranek, Cambridge, MA
Per V. Brüel, Brüel Acoustics, Denmark
William J. Cavanaugh, Cavanaugh Tocci Associates, Sudbury, MA
Floyd Dunn, University of Illinois, Urbana, IL
Ira Dyer, M.I.T. Cambridge, MA
Tony F. W. Embleton, National Research Council, Ottawa, Canada
Frank J. Fahy, I.S.V.R., Southampton University, England
James L. Flanagan, Rutgers University, Piscataway, NJ
Henning von Gierke, Wright Patterson AFB, OH
David M. Green, University of Florida, Gainesville, FL
Manfred A. Heckl,* Institute of Technical Acoustics, Berlin, Germany
Joseph E. Hind, University of Wisconsin, Madison, WI
William W. Lang, Poughkeepsie, NY
Sir James Lighthill, University College London, England
Bruce Lindsay,* Brown University, Providence, RI
Leonid M. Lyamshev, Academy of Sciences, Moscow, Russia
Richard H. Lyon, M.I.T., Cambridge, MA
Maa Dah-You, Institute of Acoustics, Beijing, P.R.C.
Philip L. Marston, Washington State University, Pullman, WA
Dennis McFadden, University of Texas, Austin, TX
Harry B. Miller, Naval Underwater Systems Center, New London, CT
Alan Powell, University of Houston, TX
Thomas D. Rossing, Northern Illinois University, DeKalb, IL
Andrew F. Seybert, University of Kentucky, Lexington, KY
Theodore J. Schultz,* Schultz & Associates, Boston, MA
W. Dixon Ward, University of Minnesota, MN

*Deceased

ENCYCLOPEDIA OF ACOUSTICS

Volume Two

MALCOLM J. CROCKER, *Editor-in-Chief*
Auburn University

A Wiley-Interscience Publication
JOHN WILEY & SONS, INC.
New York • Chichester • Weinheim • Brisbane • Singapore • Toronto

This text is printed on acid-free paper.

Copyright © 1997 by John Wiley & Sons, Inc.

All rights reserved. Published simultaneously in Canada.

Reproduction or translation of any part of this work beyond that permitted by Section 107 or 108 of the 1976 United States Copyright Act without the permission of the copyright owner is unlawful. Requests for permission or further information should be addressed to the Permissions Department, John Wiley & Sons, Inc., 605 Third Avenue, New York, NY 10158-0012

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold with the understanding that the publisher is not engaged in rendering legal, accounting, or other professional services. If legal advice or other expert assistance is required, the services of a competent professional person should be sought.

Library of Congress Cataloging in Publication Data:

Encyclopedia of acoustics / [edited by] Malcolm J. Crocker
p. cm.
Includes index.
ISBN: 0-471-18005-X (Volume Two)
ISBN: 0-471-80465-7 (set)
1. Acoustics—Encyclopedias. I. Crocker, Malcolm J.
QC221.5.E53 1997
534—dc21 96-37424

Printed in the United States of America

10 9 8 7 6 5 4

CONTENTS

Volume One

FOREWORD xv

PREFACE xvii

CONTRIBUTORS xix

PART I GENERAL LINEAR ACOUSTICS 1

1 Introduction 3
Malcolm J. Crocker

2 Mathematical Theory of Wave Propagation 21
Allan D. Pierce

3 Ray Acoustics for Fluids 39
D. E. Weston

4 Ray Acoustics for Structures 47
Leopold B. Felsen

5 Interference and Steady-State Scattering of Sound Waves 55
Herbert Überall

6 Speed of Sound in Fluids 69
Allan J. Zuckerwar

7 Standing Waves 81
U. S. Shirahatti and Malcolm J. Crocker

8 Waveguides 91
P. O. A. L. Davies

9 Steady-State Radiation from Sources 107
A. P. Dowling

10 Transient Radiation 127
Peter Stepanishen

11 Acoustic Interaction between Structures and Fluids 135
F. J. Fahy

12 **Boundary Waves** 153
Ivan Tolstoy

13 **Acoustic Lumped Elements from First Principles** 161
Robert E. Apfel

14 **Acoustic Modeling: Finite Element Method** 165
A. Craggs

15 **Acoustic Modeling: Boundary Element Methods** 173
A. F. Seybert and T. W. Wu

16 **Acoustic Modeling (Ducted-Source Systems)** 185
M. G. Prasad and Malcolm J. Crocker

PART II NONLINEAR ACOUSTICS AND CAVITATION 191

17 **Nonlinear Acoustics and Cavitation** 193
David T. Blackstock and Malcolm J. Crocker

18 **Some Model Equations of Nonlinear Acoustics** 197
David T. Blackstock

19 **Propagation of Finite-Amplitude Waves in Fluids** 203
D. G. Crighton

20 **Parameters of Nonlinearity of Acoustic Media** 219
E. Carr Everbach

21 **Finite-Amplitude Waves in Solids** 227
Mack A. Breazeale

22 **Nonlinear Standing Waves in Cavities** 237
Alan B. Coppens and Anthony A. Atchley

23 **Nonlinear Effects in Sound Beams** 249
Mark F. Hamilton

24 **Nonlinear Lumped Elements** 257
William E. Zorumski

25 **Cavitation** 263
Werner Lauterborn

26 **Sonochemistry and Sonoluminescence** 271
Kenneth S. Suslick and Lawrence A. Crum

PART III AEROACOUSTICS AND ATMOSPHERIC SOUND 283

27 **Introduction** 285
James Lighthill

28 **Aerodynamic and Jet Noise** 301
Alan Powell

29 **Interaction of Fluid Motion and Sound** 313
J. E. Ffowcs Williams

30 **Acoustic Streaming** 321
N. Riley

31 **Shock Waves, Blast Waves, and Sonic Booms** 329
Richard Raspet

32 **Atmospheric Sound Propagation** 341
Louis C. Sutherland and Gilles A. Daigle

33 **Infrasound** 367
Thomas B. Gabrielson

PART IV UNDERWATER SOUND 373

34 **Introduction** 375
Ira Dyer

35 **Essential Oceanography** 381
F. H. Fisher and P. F. Worcester

36 **Propagation of Sound in the Ocean** 391
William A. Kuperman

37 **Propagation in Marine Sediments** 409
LeRoy M. Dorman

38 **Attenuation by Forward Scattering: Measurements and Modeling** 417
Robert H. Mellen, Hans G. Schneider, and David G. Browning

39 **Volume Scattering in Underwater Acoustic Propagation** 425
Shimshon Frankenthal

40 **Backscattering from Rough Surfaces and Inhomogeneous Volumes** 441
R. P. Chapman, O. Z. Bluy, and P. C. Hines

41 **Sound Radiation from Marine Structures** 459
David Feit

42 **Transient and Steady-State Scattering and Diffraction from Underwater Targets** 469
Louis R. Dragonette and Charles F. Gaumond

43 **Quantitative Ray Methods for Scattering** 483
Philip L. Marston

44 **Target Strength of Fish** 493
Kenneth G. Foote

45 **Fundamental Underwater Noise Sources** 501
William Blake

46 **Ship and Platform Noise, Propeller Noise** 521
Robert D. Collier

47 **Underwater Explosive Sound Sources** 539
Valery K. Kedrinskii

48 **Ocean Ambient Noise** 549
Ira Dyer

49 **Sonar Systems** 559
James E. Barger

50 **Oceanographic and Navigational Instruments** 581
Robert C. Spindel

51 **Acoustic Telemetry** 591
Josko A. Catipovic

52 **Transducers** 597
Joseph E. Blue and Arnie Lee Van Buren

53 **Nonlinear Sources and Receivers** 607
Mark B. Moffett and William L. Konrad

Volume Two

PART V ULTRASONICS, QUANTUM ACOUSTICS, AND PHYSICAL EFFECTS OF SOUND 619

54 **Introduction** 621
Philip L. Marston

55 **Ultrasonic Velocity** 629
John H. Cantrell and William T. Yost

56 **Ultrasonic Relaxation Processes** 641
Henry E. Bass and F. Douglas Shields

57 **Phonons in Crystals, Quasicrystals, and Anderson Localization** 651
Julian D. Maynard

58 **Surface Waves in Solids and Ultrasonic Properties** 661
Moisés Levy and Susan C. Schneider

59 **Wave Modes in Liquid Helium** 673
Gary A. Williams

60 **Ultrasonic Instruments for Nondestructive Testing** 683
Emmanuel P. Papadakis

61 **Thermoacoustic Engines** 695
Gregory W. Swift

PART IV MECHANICAL VIBRATIONS AND SHOCK 703

62 **Introduction** 705
Manfred A. Heckl

63 **Experimental and Theoretical Studies of Vibrating Systems** 715
Paul J. Remington

64 **Vibrations of One- and Two-Dimensional Continuous Systems** 735
Manfred A. Heckl

65 **Nonlinear Vibration** 753
Lawrence N. Virgin and Earl H. Dowell

66 **Random Vibration** 765
David E. Newland

67 **Shock Analysis and Design** 785
Thomas L. Geers

68 **Acoustic Emission** 797
Kanji Ono

69 **Seismic Exploration: Acoustic Imaging for Geological Information** 811
A. J. Berkhout

70 **Effects of High-Intensity Sound on Structures** 831
B. L. Clarkson

71 **Vibration Isolation and Damping** 843
Eric E. Ungar

72 **Vibration Measurements and Instrumentation** 857
Michael Möser

73 **Machinery Condition Monitoring** 869
Govindappa Krishnappa

74 **Structure-Borne Energy Flow** 881
G. Pavic

75 **Active Vibration Control** 893
C. R. Fuller

PART VII STATISTICAL METHODS IN ACOUSTICS 909

76 **Introduction** 911
Richard H. Lyon

77 **Response Statistics of Rooms** 913
Mikio Tohyama

78 **Statistical Modeling of Vibrating Systems** 925
Jerome E. Manning

PART VIII NOISE: ITS EFFECTS AND CONTROL 937

79 **Noise Control** 939
Malcolm J. Crocker

80 **Rating Measures, Descriptors, Criteria, and Procedures for Determining Human Response to Noise** 943
Malcolm J. Crocker

81 **Hearing Protection Devices** 967
E. H. Berger and J. G. Casali

82 **Determination of Noise Emission of Sound Sources** 983
George C. Maling, Jr., and William W. Lang

83 **The Generation of Noise in Machinery, Its Control, and the Identification of Noise Sources** 991
Malcolm J. Crocker

84 **Active Noise Control** 1025
P. A. Nelson and S. J. Elliot

85 **Noise Sources and Propagation in Ducted Air Distribution Systems** 1039
Howard Kingsbury

86 **Sound Power Level Predictions for Industrial Machinery** 1049
Robert D. Bruce and Charles T. Moritz

87 **Airport Noise** 1059
Kenneth McK. Eldred

88 **Surface Transportation Noise** 1073
Robert Hickling

89 **Community Response to Environmental Noise** 1083
Sanford Fidell and Karl S. Pearsons

Volume Three

PART IX ARCHITECTURAL ACOUSTICS 1093

90 **Introduction** 1095
A. Harold Marshall and William J. Cavanaugh

91 **Sound in Enclosures** 1101
K. H. Kuttruff

92 **Sound Absorption in Enclosures** 1115
David A. Bies and Colin H. Hansen

93 **Sound Insulation: Airborne and Impact** 1129
A. C. C. Warnock and Wolfgang Fasold

94 **Ratings and Descriptors for the Building Acoustical Environment** 1161
Gregory C. Tocci

95 **Sound Propagation in Work Spaces** 1181
Ulrich J. Kurze

96 **Acoustical Guidelines for Building Design** 1189
Ewart A. Wetherill

97 **Noise Control in U.S. Building Codes** 1205
Gregory C. Tocci

98 **Noise Control for Mechanical and Ventilation Systems** 1219
R. M. Hoover and R. H. Keith

PART X ACOUSTICAL SIGNAL PROCESSING 1243

99 Introduction 1245
Andrew F. Seybert

100 Statistical Theory of Acoustic Signals 1249
Allan G. Piersol

101 Practical Considerations in Signal Processing 1261
John C. Burgess

102 Acoustic Holography 1281
Julian D. Maynard

PART XI PHYSIOLOGICAL ACOUSTICS 1291

103 Introduction 1293
Joseph E. Hind

104 Auditory Function 1301
Steven Greenberg

105 Acoustical Characteristics of the Outer Ear 1325
Edgar A. G. Shaw

106 Acoustic Properties of the Middle Ear 1337
William T. Peake and John J. Rosowski

107 Anatomy of the Cochlea and Auditory Nerve 1347
Norma B. Slepecky

108 Cochlear Mechanics and Biophysics 1357
Mario A. Ruggero and Joseph Santos-Sacchi

109 Physiology of the Auditory Nerve 1371
D. O. Kim and K. Parham

110 Anatomy of the Central Auditory Nervous System 1381
Douglas L. Oliver

111 Electrophysiology of the Central Auditory Nervous System 1389
Tom C. T. Yin and Philip H. Smith

112 Biochemistry and Pharmacology of the Auditory System 1401
William F. Sewell

113 Parallel-Hierarchical Processing of Complex Sounds for Specialized Auditory Function 1409
Nobuo Suga

PART XII PSYCHOLOGICAL ACOUSTICS 1419

114 Introduction 1421
David M. Green and Dennis McFadden

115 Auditory Masking 1427
Søren Buus

116 **Frequency Analysis and Pitch Perception** 1447
Brian C. J. Moore

117 **Functions of the Binaural System** 1461
Ervin R. Hafter and Constantine Trahiotis

118 **Loudness** 1481
Bertram Scharf

119 **Effects of High-Intensity Sound** 1497
W. Dixon Ward

120 **Clinical Audiology: An Overview** 1509
Larry E. Humes

121 **Perception of Complex Waveforms** 1521
Charles S. Watson and Gary R. Kidd

122 **Adaptation in the Auditory System** 1535
Paul J. Abbas

123 **Hearing Thresholds** 1545
William A. Yost and Mead C. Killion

Volume Four

PART XIII SPEECH COMMUNICATION 1555

124 **Introduction** 1557
J. L. Flanagan

125 **Models of Speech Production** 1565
Kenneth N. Stevens

126 **Speech Perception** 1579
Joanne L. Miller

127 **Acoustical Analysis of Speech** 1589
Gunnar Fant

128 **Techniques of Speech Coding** 1599
Bishnu S. Atal

129 **Machine Recognition of Speech** 1607
Lawrence R. Rabiner

PART XIV MUSIC AND MUSICAL ACOUSTICS 1615

130 **Introduction** 1617
Thomas D. Rossing

131 **Stringed Instruments: Bowed** 1619
J. Woodhouse

132 **Stringed Instruments: Plucked** 1627
Bernard E. Richardson

133 **Woodwind Instruments** 1635
Neville H. Fletcher

134 **Brass Instruments** 1643
J. M. Bowsher

135 **Percussion Instruments** 1653
Thomas D. Rossing

136 **Pianos and Other Stringed Keyboard Instruments** 1663
Gabriel Weinreich

137 **Pipe and Reed Organs** 1671
Jürgen Meyer

138 **Electronic and Computer Music** 1679
William Morris Hartmann

139 **Human Singing Voice** 1687
Johan Sundberg

PART XV BIOACOUSTICS 1697

140 **Introduction** 1699
Floyd Dunn

141 **Acoustical Characteristics of Biological Media** 1703
Jeffrey C. Bamber

142 **Biological Effects of Ultrasound** 1727
Edwin L. Carstensen

143 **Medical Diagnosis with Acoustics** 1739
K. Kirk Shung and Guy Cloutier

144 **Acoustical Medical Imaging Instrumentation** 1751
James F. Greenleaf

145 **Effects of Vibration and Shock on People** 1761
Henning E. von Gierke

PART XVI ANIMAL BIOACOUSTICS 1781

146 **Introduction** 1783
Arthur N. Popper

147 **Hearing and Sound Perception among Vertebrate Animals** 1789
Richard R. Fay

148 **Insect Bioacoustics** 1799
T. G. Forrest

149 **Amphibians** 1807
Randy Zelick and R. R. Capranica

150 **Bird Acoustics** 1813
Stephen Nowicki

151 **Bats and Echolocation** 1819
James A. Simmons

152 **Nonhuman Primate Vocal Communication** 1823
Marc D. Hauser

153 **Bioacoustics of Marine Vertebrates** 1831
Arthur N. Popper and Peggy L. Edds-Walton

PART XVII ACOUSTICAL MEASUREMENTS AND INSTRUMENTATION 1837

154 **Introduction** 1839
P. V. Brüel, J. Pope, and H. K. Zaveri

155 **Sound Level Meters** 1845
Robert W. Krug

156 **Sound Intensity** 1855
Malcolm J. Crocker and Finn Jacobsen

157 **Calibration of Pressure and Gradient Microphones** 1869
Victor Nedzelnitsky

PART XVIII TRANSDUCERS 1881

158 **Introduction** 1883
Harry B. Miller

159 **Transducer Principles** 1889
Elmer L. Hixson and Ilene J. Busch-Vishniac

160 **Loudspeaker Design** 1903
Bradley M. Starobin

161 **Horns** 1925
Vincent Salmon

162 **Types of Microphones** 1933
Ilene J. Busch-Vishniac and Elmer L. Hixson

163 **Public Address and Sound Reinforcement Systems** 1945
David Lloyd Klepper and Larry S. King

164 **Magnetic Recording Reproducing Systems** 1963
Marvin Camras

165 **Digital Audio** 1967
James W. Beauchamp and Robert C. Maher

166 **Hearing Aid Transducers** 1979
Mead C. Killion

REVIEWERS 1991

INDEX 1993

PART V

ULTRASONICS, QUANTUM ACOUSTICS, AND PHYSICAL EFFECTS OF SOUND

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker
Copyright © 1997 John Wiley & Sons, Inc.

54

INTRODUCTION

PHILIP L. MARSTON

1 INTRODUCTION

Mechanical radiation and vibrations of matter are of fundamental importance to various areas of physics and instrumentation. An understanding of the coupling of ultrasound with molecular, thermal, rheological, and electronic properties can facilitate an analysis of attenuation mechanisms and, in some cases, facilitate the use of ultrasound as a probe of microscopic systems and processes. Quantum theory is relevant to microscopic processes affecting the propagation of sound and also to the detection of weak mechanical vibrations. Part V contains chapters on selected physical aspects of sound and vibrations of general and contemporary interest. The present chapter includes some introductory material as well as summary descriptions of relevant physics issues not described elsewhere in Part V or elsewhere in this encyclopedia.

The scientific literature pertaining to fundamental aspects of ultrasonics and physical effects of sound is sufficiently large that it would be impossible to give a uniform coverage of areas in Part V. Consequently some references will be included in the present chapter as entry points to the literature for certain specialized topics. Monographs that are especially useful are listed in Refs. 1–5.

Special attention will be given in the present chapter to certain topics of pedagogical value relating macroscopic and quantum descriptions of vibrations. The representation of classical sound waves as a coherent superposition of phonons is described in Section 2. This serves to introduce the discussion of the theory of phonons in Chapter 57. Section 2 is also related to discussions in Sections 3 and 4 of the detection of low-amplitude vibrations and the dissipation of sound in solids. These topics are relevant to the detection of vibrations with dimensionless strain amplitudes of approximately 10^{-19}, as reviewed in Ref. 6–9 where the applications include attempts to detect gravitational radiation. These developments have furthered the understanding of quantum limitations to the measurements of small-amplitude vibrations, and certain concepts of general value will be reviewed here.

2 CLASSICAL SOUND WAVES AS A COHERENT SUPERPOSITION OF PHONONS

To model the longitudinal vibrations of a crystalline solid, consider a chain of equal masses coupled only through nearest-neighbor harmonic potentials.[5] This is, of course, an approximation to the coupling between atoms for real solids; however, it is sufficient for introducing the relationship between phonons and classical sound waves. The procedure for calculating the normal modes and frequencies for such a chain is described in Chapter 57. Associated with the kth normal mode is a normal coordinate, designated here as ζ_k, that describes the displacement amplitudes for a given mode [see Eq. (1)]. Phonons are the elementary excitations of this chain given by quantizing the uncoupled harmonic oscillator corresponding to each normal coordinate. The quantization procedure, also reviewed in Chapter 57, is described in Refs. 10 and 11 for a broader class of harmonic excitations of many-particle systems. The quantum number of the oscillator corresponding to the kth mode is a nonnegative integer $n_k = 0, 1, 2, \ldots$ and is commonly referred to as the occupation number of that mode. Raising or lowering n_k by unity corresponds to adding or removing a

---

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker
Copyright © 1997 John Wiley & Sons, Inc.

phonon from the kth mode. The energy increment per phonon is $\hbar\omega_k$ where $\hbar = 1.054 \times 10^{-34}$ J · s and ω_k is the radian frequency of the mode.

A question of general interest is how to understand classical sound waves from the phonon (quantized strain field) point of view. Since we are considering a linearized description of vibrations, the problem is analogous to the relationship between classical electromagnetic fields and photons reviewed by Scully et al.[12] Consider a chain with $N + 1$ masses m with equilibrium locations $x_j = ja$, $j = -N/2, \ldots, N/2$ where N is even and a denotes the space between masses. For definiteness, attention is restricted to the case of free ends and to symmetric modes. The contributions to the displacement from equilibrium for the jth mass due to the kth normal coordinate ζ_k are related by[10]

$$u_j = \left[\frac{2}{m(N+1)}\right]^{1/2} \sum_{k=1}^{N} \zeta_k \sin\left[\frac{\pi(2k-1)j}{N}\right], \qquad (1)$$

where the normalization is such that the dimensions of ζ_k are length·(mass)$^{1/2}$. (For the system under consideration ζ_k corresponds to a superposition of the amplitudes denoted by U_k and U_{-k} in Chapter 57.) The coordinate ζ_k is treated as the position coordinate q of a harmonic oscillator having a natural frequency ω_k corresponding to that of the kth classical mode. The conjugate momentum for the normalization used here becomes[10] $p_k = \partial L(\zeta_k, \dot{\zeta}_k)/\partial\dot{\zeta}_k = \dot{\zeta}_k \equiv d\zeta_k/dt$, where L denotes the Lagrangian for the kth mode. Recall that the wave function of a quantized harmonic oscillator[13] in the nth state is proportional to a product of the nth Hermite polynomial and a Gaussian function denoted here as $\phi_n(\zeta_k)$. The general wave function of the kth normal mode is the superposition

$$\Psi_k(\zeta_k, t) = \sum_{n=0}^{\infty} c_n \exp\left[-i\left(n + \frac{1}{2}\right)\omega_k t\right]\phi_n(\zeta_k), \qquad (2)$$

where for convenience the index k has been omitted from the occupation number n_k. The question under consideration becomes how the coefficients c_n are to be chosen such that the expectation value for the normal coordinate

$$\langle\zeta_k\rangle = \int_{-\infty}^{\infty} \Psi_k^*(\zeta_k, t)\zeta_k\Psi_k(\zeta_k, t)d\zeta_k \qquad (3)$$

reduces to that of a classical harmonic oscillator in the many-phonon limit. The coefficients are also to be chosen such that the wave packet described by $|\Psi_k|^2$ oscillates coherently in time without spreading and with a minimal uncertainty product $\Delta\zeta_k\,\Delta p_k$. From an analysis of the harmonic oscillator problem due to Schrödinger, it is known that the appropriate superposition is[12–15]

$$c_n = \frac{\alpha^n}{(n!)^{1/2}} \exp\left(-\frac{|\alpha|^2}{2}\right), \qquad (4)$$

where $\alpha = \zeta_{k0}/\sigma_k$, ζ_{k0} is the expected magnitude, and $\sigma_k = (2\hbar/\omega_k)^{1/2}$ describes the width of the packet. The wave packet is Gaussian

$$|\Psi_k|^2 = \sigma_k^{-1}\pi^{-1/2} \exp\left[-\left(\frac{\zeta_k - \langle\zeta_k\rangle}{\sigma_k}\right)^2\right], \qquad (5)$$

and Eq. (3) reduces to $\langle\zeta_k\rangle = \zeta_{k0}\cos\,\omega_k t$. See Fig. 1. The probability of n phonons in the kth mode becomes

$$|c_n|^2 = \frac{|\alpha|^{2n}}{n!} \exp(-|\alpha|^2), \qquad (6)$$

which is a Poisson distribution.[12–16] The expectation value for the number of phonons in the kth mode becomes[15]

$$\langle N_k\rangle = \sum_n n|c_n|^2 = |\alpha|^2. \qquad (7)$$

The superposition in Eqs. (2)–(6) is analogous to Glauber's description of a coherent superposition of photons giving Poisson statistics for photon counting.[15] It is noteworthy that the coherent distribution given by Eq. (6) is peaked at n near $\langle N_k\rangle$, which depends on the expected amplitude ζ_{k0}. The distribution differs greatly from the Planck distribution, which is descriptive of thermally excited photons and phonons.[12, 15, 16] Traveling classical waves in solids can be treated as an appropriately phased superposition of the standing waves considered here. For example, the traveling wave $\sin(bx - \omega t)$ may be written as $\sin(bx)\cos(\omega t) - \cos(bx)\sin(\omega t)$ where b is the wavenumber.

The uncertainty in the coordinate is usually too small to be significant in acoustics. The packet width σ_k is slightly larger than the uncertainty $\Delta\zeta_k$ and specifies where $|\Psi_k|^2$ has decreased by e^{-1}. Associated with σ_k is a coordinate spread at an antinode that from Eq. (1) becomes

$$\sigma_u = \left[\frac{2}{m(N+1)}\right]^{1/2} \sigma_k \approx 2\left(\frac{\hbar}{\omega_k M}\right)^{1/2}, \qquad (8)$$

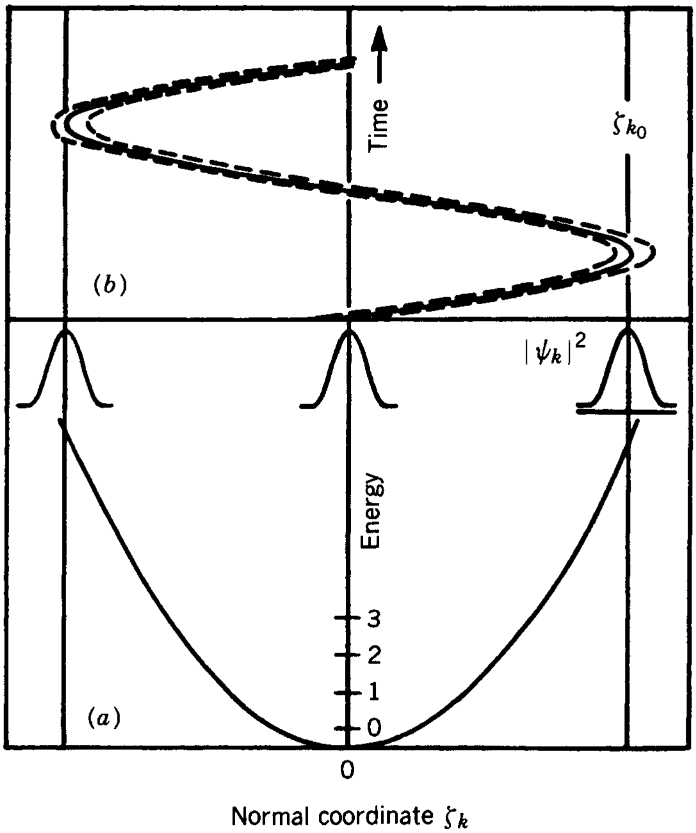


Fig. 1 Coherent state phonon description of harmonic sound waves in a crystal lattice: (*a*) shows the potential as a function of the normal coordinate ζ_k that specifies the strain amplitude for the kth normal mode. The lowest quantized energy levels are indicated corresponding to occupation numbers n of 0, 1, 2, and 3. Classical sound waves correspond to a coherent superposition of number states, specified by Eqs. (2) and (4), that give the harmonically oscillating Gaussian wave packet illustrated. The oscillating expectation value $\langle\zeta_k\rangle$ is indicated by the solid curve in (*b*). Associated with the width of that wave packet is a spread of ζ_k indicated by the dashed lines in (*b*).

where $m(N+1)$ is the mass M of the system. This has the same magnitude as the position uncertainty[6] for a quantized oscillator of mass M and frequency ω_k. For the purpose of illustration, consider the oscillations of a large crystal having $M = 1$ kg and a fundamental frequency $\omega_k/2\pi = 30$ kHz. These parameters are taken to correspond roughly to the sapphire rods described in Ref. 6 that are of interest because of their high mechanical quality factor Q (see Section 4). The coordinate spread from Eq. (8) is $\sigma_u \approx 5 \times 10^{-20}$ m. The estimated small magnitude of the uncertainty illustrates why the quantum uncertainty is generally of little importance in acoustics.

One essential character of a classical sound wave, the harmonic oscillations of $\langle\zeta_k\rangle$, can be preserved with a simpler superposition of phonon number states. Only two values of c_n having $n \gg 1$ need be nonvanishing, as described in Chapter 57. For that superposition, however, the width of the oscillating wave packet does not remain constant as in the case of the coherent state illustrated in Fig. 1.

3 DETECTION OF LOW-AMPLITUDE VIBRATIONS: SOME PHYSICAL LIMITATIONS

3.1 Thermal and Quantum Limitations

The physical limitations to the detection of the response of a mechanical resonator to weak applied forces will now be summarized. This will extend the discussion of uncertainties given in conjunction with Eq. (8) to other limitations. Figure 2 shows a diagram of the essential aspects of the detection problem. The *displacement* response of the resonator is converted to an electrical signal by a transducer and the signal is amplified and filtered. The electrical components are modeled as ideal

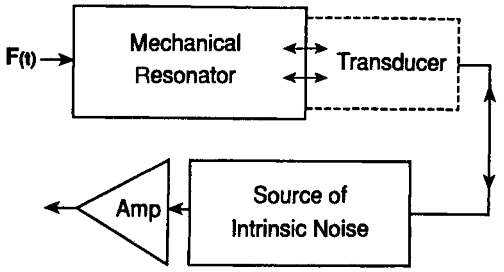


Fig. 2 System for detecting vibrations in response to a weak applied force that may be distributed over the resonator to couple into the mode of interest. With appropriate cooling and mechanical isolation of the resonator, vibration detection may be quantum mechanically limited. With a weakly coupled electrical transducer, modifications of this system may be used to study the intrinsic quality factor Q for the normal mode as described in Ref. 6.

with their inherent electrical noise introduced separately as indicated in the diagram. At low signal levels the flow of energy from the noise source back through the transducer and into the resonator becomes important. The applied force $\mathbf{F}(t)$ may represent, for example, the coupling to an external acoustic field or other influences such as gravitational radiation[6–9] or probe forces as in atomic force microscopy.[17] The summary below relies on discussions in Refs. 6–9 and on standard results for the Brownian motion of oscillators.

Consider the case of a resonator with free ends. A convenient measure of the mode amplitude is the surface displacement x measured by a transducer positioned at a displacement antinode. [For comparison with Section 2, x scales with the normal coordinate ζ_k by a factor of order $(2/M_k)^{1/2}$.] The equipartition theorem of statistical mechanics gives the mean-squared displacement from the thermal energy expression

$$\tfrac{1}{2}M_k\omega_k^2\langle x^2\rangle_{\mathrm{th}} = \tfrac{1}{2}k_BT, \tag{9}$$

where k_B denotes Boltzman's constant, T the absolute temperature of the resonator, and M_k and ω_k are the effective mass and frequency of the resonator for the mode under consideration. Typically the mode with the lowest ω_k is the one of interest. In Eq. (9), $M_k\omega_k^2$ represents the effective stiffness of the resonator, and this result has been used to estimate the thermal noise limitation of fiberoptic sensors.[18] The root-mean-square (rms) noise displacement is seen to be proportional to $T^{1/2}$. The resonator is taken to have a quality factor $Q \gg 1$ for the mode under consideration; this is the number of radians of oscillation for the mode energy to be reduced by a factor e^{-1}. The amplitude decay time is $\tau_k = 2Q/\omega_k$. The time required for thermal motion to cause a change in displacement having the same magnitude as the rms thermal displacement is determined by τ_k. Suppose the transducer samples x with a sampling time $\tau_s \le \pi/\omega_k$. During an interval τ_s the thermal change in displacement Δx_{th} is limited in magnitude by approximately[7]

$$|\Delta x_{\mathrm{th}}|^2 \approx 2\,\frac{\tau_s}{\tau_k}\,\langle x^2\rangle_{\mathrm{th}} = \frac{\tau_s\omega_k}{Q}\,\langle x^2\rangle_{\mathrm{th}} \tag{10}$$

for $\tau_s \ll \tau_k$ and factors of order unity neglected. The important point is that if the displacements are only due to thermal motion, the successive displacement samples are correlated over a range that is much smaller than the rms displacement from Eq. (9). From Eq. (10), the magnitude of displacement changes in response to suitable externally applied forces (such as impulsive forces) is limited in principle by[6,7]

$$|\Delta x_{\mathrm{th}}| \approx \left[\frac{k_BT\tau_s}{M_k\omega_kQ}\right]^{1/2} \tag{11}$$

instead of the rms displacement from Eq. (9). Comparison with Eq. (9) gives an effective noise temperature that can be much less than the actual temperature T when Q is large and narrow-band sensitivity is desired. In practice, the sensitivity may also be limited by amplifier noise and the coupling of that noise back to the resonator.[6–8]

The discussion above is sufficient to understand why resonators with large Q's have been developed for the purpose of detecting responses to weak applied forces. It is reasonable to examine how large Q needs to be before the limiting sensitivity is determined by quantum mechanical uncertainty instead of by thermal excitations. The estimated threshold value is[6] $Q_{\mathrm{qm}} \approx k_BT\tau_s/\hbar$, which is in agreement with the magnitude estimated by comparing Eqs. (8) and (11). Cryogenic resonators have been constructed having quality factors in excess of Q_{qm} for sampling times $\approx 10^{-4}$–10^{-3} s, indicating that quantum mechanical considerations may be important for some mechanical resonators.[6] (See the discussion in Section 4 concerning limitations on the Q of resonators.)

3.2 Squeezed States of Mechanical Resonators and Phonons

Methods for extending the limits of detectability beyond those for an oscillating wave packet in the coherent state described by Eqs. (2)–(7) have been proposed by various authors. The experimental emphasis has been on improving the measurability of selected features of light waves by superposing photon states in a way that differs from the classical coherent state described in Section 2.[19] Suppose that it is desired to improve the quantum limitations on the measurement of amplitude beyond those implied by the optical version of Fig. 1 in which the normal coordinate ζ_k is replaced by the instantaneous electric field amplitude. The coefficients c_n of the superposition in Eq. (2) are changed to reduce the amplitude uncertainty with a corresponding increase in the momentum uncertainty Δp_k. The resulting nonclassical states of light are referred to as squeezed states.[19] The squeezing of the quantized strain of a mechanical resonator has been proposed as a method of extending the quantum limit of detectability of weak vibrations.[9] Theoretical aspects of squeezed phonon states have also been discussed. Squeezing is closely related to the concept of quantum nondemolition measurements for evading or altering the usual consequences of measurement for a quantized oscillator.[7, 19] The interest in squeezing uncertainties is not limited to quantum noise. Thermal noise (Brownian motion) of a mechanical resonator has been reduced by a form of squeezing.[17] The resonator was a cantilever with a flexural normal mode.

4 INTRINSIC LOSS MECHANISMS AND THE ATTENUATION OF SOUND IN SOLIDS AND FLUIDS

While the attenuation of sound is a phenomenon of general interest, the mechanisms of attenuation also limit the quality factor Q of the mechanical resonators mentioned in Section 3. Generally for a perfect nonmetallic crystal, the most important attenuation mechanisms for longitudinal waves are thermoelastic dissipation and phonon–phonon interactions.[2,3,5,6] Thermoelastic dissipation is the consequence of heat flow associated with local temperature changes resulting from the oscillating compression and expansion of the materials. (The analogous attenuation mechanism for fluids is discussed in Chapter 56). Phonon–phonon interactions are a consequence of the anharmonicity of the crystal lattice that is not included in the phonon formulations of Section 2 and Chapter 57. In addition to the sound wave of interest, thermal phonons will be present with a Planck distribution of energies and typical frequencies of 10^{12} Hz. Assuming that the acoustic wavelength is much longer than the mean free path of the thermal phonons, the lattice nonlinearity results in a shift of the phonon frequencies. The return of the phonon distribution to thermal equilibrium contributes to the attenuation of sound as originally pointed out by Akhiezer.[2,3,6] The resulting attenuation generally increases in proportion to ω^2 where ω is the frequency of the sound wave. An important intrinsic mechanism for the attenuation of sound in metals is the interaction of electrons and phonons, which is discussed in Chapter 58. The aforementioned mechanisms are thought to be the important ones for limiting the Q of mechanical resonators at moderate to low temperatures. Generally Q tends to increase with decreasing temperature T. For frequencies on the order of 10^4 Hz and $T \approx 4$ K, Q values of 5×10^9 and 2×10^9 have been achieved for single crystal bars of sapphire and silicon.[6,7] Engineering solids exhibit a broad range of attenuation mechanisms. These include lattice defects and mechanisms resulting from the polycrystalline internal structure.[2,3,6] Various specialized interactions leading to attenuation of sound and transverse waves in solids have also been studied.[1-3]

The attenuation of sound in gases and liquids is generally dominated by viscous, thermal, and relaxation mechanisms as explained in Chapter 56. Ordinary sound waves in liquid helium, discussed in Chapter 59, are noteworthy because of the combination of low attenuation and low sound speed.

5 BALLISTIC PHONONS, PARTICLE DETECTION, AND PICOSECOND ULTRASONICS

In nonmetallic crystals, the principal source of thermal conduction is thermal phonons having typical frequencies of 10^{11}–10^{12} Hz. Measurements for crystals of the thermal conductivity tensor for temperatures T below about 20 K show that conduction tends to decrease as T is lowered. This indicates that the phonon mean free path has grown to the dimensions of the crystal and is limited by boundary scattering. Consequently for sufficiently low T, heat energy that ordinarily propagates diffusively as a consequence of internal phonon scattering can be carried ballistically across the crystal without scattering.[16] This phenomenon has a number of potential applications and has enabled studies of elastic properties at ultrahigh frequencies. Typically the ballistic phonons are produced on the surface of the crystal as a burst of nonequilibrium phonons excited by a focused laser beam. Time-resolved arrivals of groups of phonons associated with the different classes of elastic waves (longitudinal, fast transverse, and slow transverse) are detected with high-speed bolometers typically made of superconducting thin film. Since the phonons are localized in both space and time, they consist of a superposition of the occupation number states for several normal modes of the lattice.

An important property of ballistic phonons is that as a consequence of propagation in nonspecific directions of the anisotropic crystal, the phonon flux is focused along caustic directions. This follows from the usual anisotropic properties of crystals at ordinary ultrasonic frequencies.[20] The flux pattern at the surface may in principle be inverted to determine the location of the transient heat source. A promising application is the detection of nuclear radiation incident on silicon crystals.[21] Sensitivity has been demonstrated to α particles and x-rays.

A development that has extended the frequency range of room temperature ultrasonic measurements is linked with the development of high-power short-pulse lasers. The availability of intense laser sources having durations of 0.1–1 ps has facilitated the photoacoustic generation of transient stress waves having duration ~10 ps and frequencies in the 100-GHz range.[22] The duration of the waves is determined by the intrinsic material response rather than the duration of the optical pulse. The reflectivity of short lower power optical pulses, delayed relative to the initial pulse, are used to monitor the reflectivity and propagation of acoustic transients. At room temperature the technique is primarily useful for investigating the propagation across thin-layered structures.

6 RADIATION PRESSURE, ACOUSTIC LEVITATION, AND LOW-FREQUENCY OSCILLATIONS DRIVEN BY MODULATED ULTRASOUND

The radiation pressure on objects placed in propagating and standing acoustic waves in fluids is an effect of sound that has diverse applications both within and outside the field of acoustics. For example, the average force on liquid or elastic spheres can be used to measure the acoustic amplitude.[23,24] It is noteworthy, however, that the radiation pressure produced by a beam of sound is affected by the boundary conditions on the fluid.[24,25] Several applications of radiation pressure make use of the ability to levitate a fluid or solid object without mechanical contact. Containerless acoustic trapping of solid, liquid, and molten materials is carried out with ultrasonic standing waves in air (or other gases) where the weight of the object is supported by the radiation pressure.[26] The method has been used for measuring the properties of materials. For trapping objects acoustically in microgravity (such as in NASA's Space Shuttle) sound in the low kilohertz range has been used.[27] Acoustic trapping of tiny bubbles in water has been used to facilitate controlled studies of radial bubble pulsations (see Chapters 25 and 26). The radiation force on a bubble in a standing wave is sometimes termed the primary Bjerknes force. There can also be a radiation force due to the sound scattered by a neighboring bubble, which is usually attractive, and is termed the secondary Bjerknes force.[28]

Low-frequency oscillations of levitated fluid objects (drops or bubbles) can be excited by temporally modulating either the levitation standing wave or a second auxiliary standing wave introduced to control the shape of an object. This has been demonstrated for oil drops[27,29] and large bubbles[30] levitated in water and drops of water and other liquids levitated in air. The excitation of such low-frequency oscillations is noteworthy in that though the coupling is second order in the acoustic pressure amplitude as with other aspects of radiation pressure, the response of the object need not be small when the modulation is tuned to one of the resonances. Capillarity typically provides the restoring force. There is a low-frequency flow induced by the oscillations that is distinct from the acoustic streaming normally associated with attenuation as described in Chapter 30. Sound can also be used to apply controlled radiation torques on levitated objects.[27]

7 DIFFRACTION OF LIGHT BY SOUND: ACOUSTOOPTICS

Traveling or standing acoustical waves affect light because of the variations of the optical index of refraction that accompany the density variations in the sound wave. This acoustooptic effect has been used to measure the wavelength of sound and is briefly discussed in Chapter 55. Other important applications include the modulation and control of light[31,32] and the imaging of acoustic wavefields by schlieren methods.[33] A compilation of several of the original studies has been published,[31] and Ref. 1 gives reviews of the various applications. Strain-induced modulations of the refractive index are also relevant to fiberoptic hydrophones discussed in Refs. 18 and 32 and in Chapter 159.

8 GENERATION OF SOUND WITH MODULATED LASER LIGHT: PHOTOACOUSTICS

The absorption of modulated laser light by solids, liquids, and gases results in a time-dependent thermal expansion that is accompanied by the radiation of sound. The rate of thermalization and expansion is such that the acoustic frequencies attainable are limited to about 100 GHz as noted in Section 5. Most applications are at much lower frequencies and the local heating rate is simply proportional to $\alpha I(\mathbf{r}_s, t)$ where α is the absorption coefficient, I is the space and time-dependent optical irradiance in watts/centimetre square, and $\mathbf{r}_s$ is a source point location under consideration. The temporal profile of the acoustic pressure depends on the spatial properties of the source and the position of the observer.[32] For example, consider an N-dimensional source with $N = 1, 2$, and 3 corresponding to absorption confined to a thin fluid layer, thin cylinder, and small droplet, respectively. In the absence of any boundary reflections, the corresponding pressures for the low-frequency spectral components $p(\mathbf{r}_s, t')$ evaluated at time t' retarded by the propagation delay, become proportional to $I(\mathbf{r}_s, t)$, $(d/dt)^{1/2} I(\mathbf{r}_s, t)$, and $dI(\mathbf{r}_s, t)/dt$, respectively, as reviewed by Diebold et al.[34] Here $(d/dt)^{1/2}$ is the half-order derivative operator. For the case of a distant observer and a compact source ($N = 3$), the singularity in p implied by dI/dt for a step change in I is softened, indicating that the relationships noted above need to be corrected for many real situations. Furthermore, for large I the generation process becomes nonlinear and can be complicated by boiling.[32]

Photoacoustic sources have numerous applications in materials characterization[35] and nondestructive evaluation[36] where they have the advantages of a noncontact source. See also reviews in Ref. 1. Photoacoustics is useful for measuring weak optical absorption coefficients, and the acoustical signal can be enhanced by acoustical resonances of the absorber. As noted in Chapter 56, photoacoustic sources can be useful for investigating relaxation processes affecting the attenuation of sound.

9 SURFACE ACOUSTIC WAVE DEVICES AND ULTRASONIC INSTRUMENTATION

Chapter 58 reviews the role of surface acoustic waves (SAW) in the investigation of various microscopic processes, though SAW devices have much broader applications. They are extensively used in real-time filtering and signal processing with typical operating frequencies in the range from 10 MHz to 1 GHz.[37] Applications include the sensing of a wide range of physical properties.[1] Chapter 60 concerns ultrasonic instrumentation for nondestructive testing and evaluation. Relevant principles are discussed in Ref. 37.

10 THERMOACOUSTIC ENGINES AND REFRIGERATORS

Chapter 61 describes a relatively recent development. Reliable refrigerators have been demonstrated that are driven by sound waves.[38] One of the advantages is that the working fluids can be environmentally safer than conventional refrigerants. A reciprocal process where sound is generated by temperature differences has also been developed.

Acknowledgment

Preparation of this chapter was supported by the U.S. Office of Naval Research.

REFERENCES

1. W. P. Mason and R. N. Thurston or R. N. Thurston and A. D. Pierce (Eds.), *Physical Acoustics*, Vols. 1–23, Academic, New York. (Note: Current volumes typically list the chapter titles of previous volumes.)
2. R. T. Beyer and S. V. Letcher, *Physical Ultrasonics*, Academic, New York, 1969.
3. R. Truell, C. Elbaum, and B. B. Chick, *Ultrasonic Methods in Solid State Physics*, Academic, New York, 1969.
4. P. D. Edmonds (Ed.), *Ultrasonics, Methods of Experimental Physics*, Vol. 19, Academic, New York, 1981.
5. C. Kittel, *Introduction to Solid State Physics*, 6th ed., Wiley, New York, 1986.
6. V. B. Braginsky, V. P. Mitrofanov, and V. I. Panov, *Systems with Small Dissipation*, University of Chicago Press, Chicago, 1985; M. F. Bocko and R. Onofrio, "On the Measurement of a Weak Classical Force Coupled to a Harmonic Oscillator: Experimental Progress," *Rev. Mod. Phys.*, Vol. 68, 1996, pp. 755–799.
7. K. S. Thorne, "Gravitational Radiation," in S. W. Hawking and W. Israel (Eds.), *Three Hundred Years of Gravitation*, Cambridge University Press, Cambridge, 1987, pp. 330–458.
8. P. F. Michelson, et al., "Near Zero: Toward a Quantum-limited Resonant-Mass Gravitational Radiation Detector," in J. D. Fairbank et al. (Eds.), *Near Zero: New Frontiers of Physics*, Freeman, New York, 1988, pp. 713–730.
9. J. N. Hollenhorst, "Quantum Limits on Resonant-Mass Gravitational-Radiation Detectors," *Phys. Rev. D*, Vol. 19, 1979, pp. 1669–1679.
10. A. L. Fetter and J. D. Walecka, *Theoretical Mechanics of Particles and Continua*, McGraw-Hill, New York, 1980, Chaps. 4 and 6 and Problem 4.15.
11. A. L. Fetter and J. D. Walecka, *Quantum Theory of Many-Particle Systems*, McGraw-Hill, New York, 1971.
12. M. O. Scully and S. F. Jacobs, "Coherence—A Sticky Subject," *Appl. Opt.*, Vol. 9, 1970, pp. 2414–2422; M. O. Scully and M. Sargent, "The Concept of the Photon," *Phys. Today*, Vol. 25, No. 3, 1972, pp. 38–47.
13. L. I. Schiff, *Quantum Mechanics*, 3rd ed., McGraw-Hill, New York, 1968, pp. 66–76.
14. E. Goldin, *Waves and Photons*, Wiley, New York, 1982, Chaps. 7 and 8.
15. R. Loudon, *The Quantum Theory of Light*, University Press, Oxford, 1983, Chap. 4.
16. J. P. Wolfe, "Ballistic Heat Pulses in Crystals," *Phys. Today*, Vol. 33, No. 12, 1980, pp. 44–50.

17. D. Rugar and P. Grutter, "Mechanical Parametric Amplification and Thermomechanical Noise Squeezing," *Phys. Rev. Lett.*, Vol. 67, 1991, pp. 699–702.
18. T. J. Hofler and S. L. Garrett, "Thermal Noise in a Fiber Optic Sensor," *J. Acoust. Soc. Am.*, Vol. 84, 1988, pp. 471–475.
19. M. C. Teich and B. E. A. Saleh, "Squeezed and Antibunched Light," *Phys. Today*, Vol. 43, No. 6, 1990, pp. 26–34; P. Meystre and D. F. Walls (Eds.), *Nonclassical Effects in Quantum Optics*, American Institue of Physics, New York, 1991.
20. M. J. P. Musgrave, *Crystal Acoustics*, Holden-Day, San Francisco, 1970.
21. B. A. Young, B. Cabrera, and A. T. Lee, "Observations of Ballistic Phonons in Silicon Crystals Induced by α Particles," *Phys. Rev. Lett.*, Vol. 64, 1990, pp. 2795–2798.
22. H. T. Grahn, H. J. Maris, and J. Tauc, "Picosecond Ultrasonics," *IEEE J. Quant. Electronics*, Vol. 25, 1989, pp. 2562–2569.
23. T. Hasegawa, "Comparison of Two Solutions for Acoustic Radiation Pressure on a Sphere," *J. Acoust. Soc. Am.*, Vol. 61, 1977, pp. 1445–1448.
24. R. T. Beyer, "Radiation Pressure—The History of Mislabeled Tensor," *J. Acoust. Soc. Am.*, Vol. 63, 1978, pp. 1025–1030.
25. C. P. Lee and T. G. Wang, "Acoustic Radiation Pressure," *J. Acoust. Soc. Am.*, Vol. 94, 1993, pp. 1099–1109.
26. E. H. Trinh, "Compact Acoustic Levitation Device for Studies in Fluid Dynamics and Material Science in the Laboratory and Microgravity," *Rev. Sci. Inst.*, Vol. 56, 1985, pp. 2059–2065.
27. T. G. Wang, "Equilibrium Shapes of Rotating Spheroids and Drop Shape Oscillations," *Adv. Appl. Mech.*, Vol. 26, 1988, pp. 1–26.
28. L. A. Crum, "Bjerknes Forces on Bubbles in a Stationary Sound Field," *J. Acoust. Soc. Am.*, Vol. 57, 1975, pp. 1363–1370.
29. P. L. Marston and R. E. Apfel, "Quadrupole Resonance of Drops Driven by Modulated Acoustic Radiation Pressure—Experimental Properties," *J. Acoust. Soc. Am.*, Vol. 67, 1980, pp. 27–37.
30. T. J. Asaki, P. L. Marston, and E. H. Trinh, "Shape Oscillations of Bubbles in Water Driven by Modulated Ultrasonic Radiation Pressure: Observations and Detection with Scattered Laser Light," *J. Acoust. Soc. Am.*, Vol. 93, 1993, pp. 706–713.
31. A. Korpel (Ed.), *Selected Papers on Acousto-Optics*, SPIE, Bellingham, WA, 1990.
32. L. M. Lyamshev, "Lasers in Acoustics," *Sov. Phys. Usp.*, Vol. 30, 1987, pp. 252–279.
33. S. Stanic, "Quantitative Schlieren Visualization," *Appl. Opt.*, Vol. 17, 1978, pp. 837–842.
34. G. J. Diebold, T. Sun, and M. I. Khan, "Photoacoustic Monopole Radiation in One, Two, and Three Dimensions," *Phys. Rev. Lett.*, Vol. 67, 1991, pp. 3384–3387.
35. D. A. Hutchins and A. C. Tam, "Pulsed Photoacoustic Materials Characterization," *IEEE Trans. Ultrasonics, Ferro., Freg. Cont.*, Vol. UFFC-30, 1986, pp. 429–449 (and other articles in this special issue).
36. F. A. McDonald, "Practical Quantitative Theory of Photoacoustic Pulse Generation," *Appl. Phys. Lett.*, Vol. 54, 1989, pp. 1504–1506.
37. G. S. Kino, *Acoustic Waves: Devices Imaging, and Analog Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1987.
38. G. W. Swift, "Thermoacoustic Engines and Refrigerators," *Phys. Today*, Vol. 48, No. 7, 1995, pp. 22–28.

55

ULTRASONIC VELOCITY

JOHN H. CANTRELL AND WILLIAM T. YOST

1 INTRODUCTION

Measurements of the sound velocity and attenuation can provide considerable information about the physical properties of solids, liquids, and gases. Measurement of the variation of the sound velocity as a function of some intrinsic material variable, such as temperature or pressure, provides additional information about basic (often nonlinear) material properties. It is important to distinguish in measurements of velocity the difference between group and phase velocities. This distinction is addressed below in the context of acoustic propagation in a dispersive medium. Similarly, it is important to distinguish between attenuation and absorption in acoustic measurements. Absorption refers to the loss of amplitude of an acoustic wave that results in an increase of temperature (however slight) in the propagation medium. Attenuation refers to the total change in the amplitude of an acoustic wave resulting from all mechanisms responsible for such changes, including but not limited to absorption, scattering, diffraction, and phase cancellation.

Numerous techniques have been reported in the literature for making velocity and attenuation measurements. A number of review articles have been published emphasizing different applications or aspects of the measurement techniques.[1-6] The most common techniques are based on optical, pulse, and continuous-wave methods or hybrids of these methods. The particular technique chosen is dictated in part by the availability of equipment and ease of set-up for the particular application, the specific material and ultrasonic properties of the material to be measured, and the accuracy and precision desired in the measurement.

This chapter is mainly concerned with an examination of representation measurement techniques in each of the aforementioned categories. Pulse methods are slightly emphasized because of their widespread popularity and ease of application. The accuracy and precision of a given measurement are intrinsically related to error sources and are among the most important considerations in evaluating data from any technique. The chapter concludes with a general discussion of the most common, and often most correctable, sources of error.

2 PULSE SYSTEMS

2.1 Basic Pulse–Echo Method

An equipment arrangement for the basic pulse–echo system[7,8] that allows for both velocity and attenuation measurements is shown in Fig. 1. For velocity measurements all switches are set to the "out" position and a pulsed ultrasonic signal is transmitted into the sample by means of a transducer attached to the sample surface. The ultrasonic pulse travels through the sample and reflects between the sample boundaries, eventually decaying away because of sample attenuation. Each time the ultrasonic pulse strikes the end of the sample coupled to the transducer, an electrical signal is generated that is amplified and displayed on an oscilloscope. The received signals are commonly rectified and filtered (i.e., detected) before being displayed on the oscilloscope. The velocity of ultrasonic wave propagation is determined by measuring the transit time between corresponding reference points of the reflected pulses and the corresponding pulse propagation distance in the sample.

Adjusting the attenuator to maintain a constant receiver signal level for each displayed echo minimizes

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker
Copyright © 1997 John Wiley & Sons, Inc.

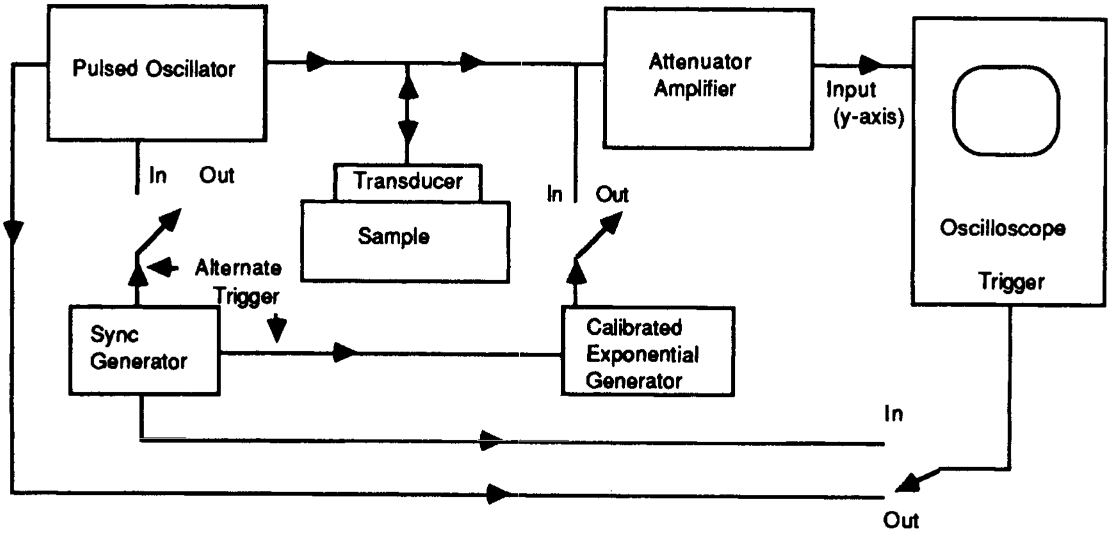


Fig. 1 Basic pulse–echo system (switches "out") for velocity and attenuation measurements and improvements (switches "in") for attenuation measurements.

the error due to amplifier nonlinearities and reduces the uncertainty in the signal reference points. The ultrasonic attenuation can also be read directly from the calibrated attenuator with this procedure.

A small transit time error is introduced into velocity measurements of solids using contact transducers in part because of the partial transmission and reflection of the ultrasonic pulse at the sample–transducer interface. This phenomenon changes the shape of the pulse with each successive echo and foils any attempt to find some characteristic "mark" on the pulses to use as an accurate reference point to measure transit time. Reference mark or echo-matching measurements,[9] while expedient, can lead to errors as large as a few parts in 10^3. Cycle-for-cycle matching of the echoes[10] eliminates this particular aspect of the problem and is discussed in Section 2.3. Ultrasonic phase shifts incurred upon pulse reflection at the sample–transducer interface are another source of error that also affect cycle-for-cycle matching and are addressed in Section 2.2.

In measurements of solids using contact transducers, the basic pulse–echo method using leading-edge echo matching is generally accurate to a few parts in 10^3. Attenuation measurements accurate to about 5 parts in 10^2 are possible. In liquids and gases the sound velocity can be determined by measuring the distance the transducer must be moved to delay the received echo by a specified amount. This variable-path-length procedure increases the accuracy of the basic pulse–echo method to about 5 parts in 10^4, since multiple reflections need not be considered.

The accuracy of attenuation measurements can be improved by replacing the calibrated attenuator with either a calibrated exponential generator or a pulser having a calibrated attenuator. This arrangement corresponds to setting all switches to the "in" position in Fig. 1. In the pulser arrangement,[11] measurements are made by matching the comparison pulse amplitude to the selected echo on alternate sweeps of the oscilloscope. Accuracies of the order 2 parts in 10^2 are possible with this set-up.[12] If the exponential generator is used,[13] the generated curve is matched to a selected pair of peaks of the decaying echo train. In both cases correction for diffraction losses is made for the echoes selected. Automated systems using peak detection of two selected echoes have been developed that achieve greater sensitivities than the manual systems.[5]

Finally, a modification useful for velocity measurements may be made in the basic pulse–echo method by placing a second transducer (used as a receiver) at the end of the sample opposite the transmitting transducer.[14] The received signal is used to retrigger the pulse generator, thereby generating a continuous succession of pulses. The repetition rate of the pulse sequence may be used to determine the ultrasonic velocity, since the repetition rate depends on the pulse travel time in the medium.

An inherent error in the pulse repetition rate (due to time delays in the electronic circuits, changes in pulse shape from attenuation, and presence of transducer–sample interfaces) limits the accuracy of this sing-around method for absolute velocity measurements. Accuracies of a few parts in 10^4 for absolute velocity measurements are possible with this method. Improved versions of the sing-around method[15] allow relative velocity measurements (i.e., changes in velocity) to be measured with accuracies to 1 part in 10^7.

2.2 Gated Double-Pulse Superposition Method

This method[16] utilizes a pulse generated by gating a continuously running oscillator. The gated, continuous, ultrasonic wave (toneburst) is transmitted into the sample followed by a second ultrasonic toneburst phase locked but delayed in time with respect to the first. The two tonebursts reflect between the ends of the sample, giving rise to two pulse–echo trains. The time delay is adjusted such that superposition of the desired echoes from the two pulse trains is achieved. The resulting signal is received by a transducer at the opposite end of the sample, amplified, and then displayed on an oscilloscope.

The superimposed signal may be made zero, independently of time, by adjusting the oscillator frequency $f = \omega/2\pi$ such that the time delay τ between the pulses and the phase shift γ due to reflection at the sample ends satisfy the condition[17] $\omega\tau - \gamma = (2n+1)\pi/2$, where n is a nonnegative integer. The pulse transit time is then calculated from this expression after n and γ have been determined. The phase angle γ is given in Ref. 17 in terms of the acoustical characteristic impedances Z_B, Z_T, and Z_S (i.e., the product of the mass density and the sound velocity) of the bonding material, transducer, and sample, respectively, and the thicknesses and velocities of sound of the transducer and the bonding material.

If the bond thickness can be neglected, the expression for γ is greatly simplified. For noncontact transducers, $Z_T = 0$ and the transit time is simply calculated as $\tau = 1/2(f_{n+1} - f_n)$. Accuracies of 1 part in 10^4 are possible with the gated double-pulse superposition method.[16]

2.3 Pulse Superposition Method

The equipment arrangement for the pulse superposition method[18,19] is shown in Fig. 2. A series of pulses (tonebursts) from a radio frequency (RF) pulse generator is introduced into the sample. The repetition rate of these pulses, controlled by the frequency of a continuous-wave (CW) low-frequency oscillator is adjusted to correspond approximately to some multiple p of an acoustic round-trip transit time δ ($=2\tau$) in the sample. The oscillator frequency is adjusted to produce an "in-phase" superposition of pulses such that a maximum amplitude is achieved. The measured time delay T between superimposed pulses is the reciprocal of the CW oscillator frequency and is written as[19]

$$T = p\delta - \frac{p\gamma}{360f} + \frac{n}{f}, \qquad (1)$$

where γ again is the phase angle associated with the pulse reflections, f is the ultrasonic frequency, and n is now a positive or negative integer that indicates the cyclic mismatch between pulses.

The pulse transit time τ is calculated from Eq. (1) once a value of T corresponding to a given n is determined. Procedures for obtaining the appropriate value for T is given in Ref. 19. For measurements in solids using bonded-contact transducers the unknown bond thickness introduces some uncertainty in the values of the γ's. In this case the phase angles are calculated as a function of estimated bond thickness, and these values are used to calculate the change in T (i.e., ΔT) as a function of the bond thickness. Such calculations establish that ΔT has a limited range for a given value of p and n. Thus all experimentally measured values of T and ΔT outside the limits calculated can be eliminated from consideration.[19]

Absolute velocity measurements having accuracies of 2 parts in 10^4 have been reported with the pulse superposition method.[18,19] Transit time measurements, hence relative velocity measurements, can be made to 1 part in 10^7.

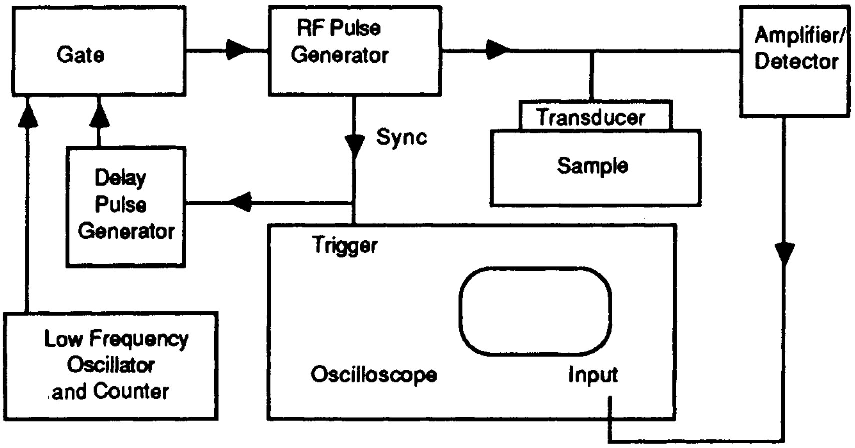


Fig. 2 Equipment arrangement for pulse superposition method.[18,19]

2.4 Echo-Overlap Method

A schematic of the equipment arrangement for the echo-overlap method[9,20,21] is presented in Fig. 3. A series of RF pulses from an RF-pulsed oscillator is transmitted into the sample. The trigger signal is derived by frequency dividing the output of a CW low-frequency (typically 100–1000 kHz) oscillator by a factor of 10^3. The trigger is used to control the pulse repetition rate and to trigger a double-time-delay (strobe) generator that actuates the Z-axis intensity gate on the oscilloscope. The delays are adjusted to intensify any chosen pair of displayed echoes. The oscilloscope is operated with a linear sweep (trigger-sweep generator position) during this adjustment so that many echoes appear on the screen.

During the acoustic transit time measurement the oscilloscope is switched to an X–Y mode of operation. In this mode the low-frequency CW oscillator provides the sweep without decade division and the oscilloscope intensity is reduced so that only the two intensified echoes of interest are visible. If the CW oscillator frequency is adjusted to correspond to the reciprocal of the travel time between the echoes, the echoes can be made to overlap cycle for cycle. Overlap also occurs for an integral multiple m of this frequency, since an echo then appears for every mth sweep. Centering of the overlap pattern can be accomplished with the phase shift network. The round-trip acoustic transit time δ ($=2\tau$) in the sample is obtained from Eq. (1), where the time delay actually measured between the echoes is the product of m and the reciprocal of the oscillator frequency. The procedure for obtaining the correct cycle-for-cycle matching of the chosen echo pair (i.e., $n = 0$ condition) and the calculation of the phase shift due to reflection are outlined in Refs. 18 and 19.

Absolute velocity measurements of a few parts in 10^5 have been reported with the echo-overlap method.[9]

2.5 Pulse Interferometer Methods

In the pulse interferometer methods an ultrasonic toneburst is transmitted into the sample by gating a CW

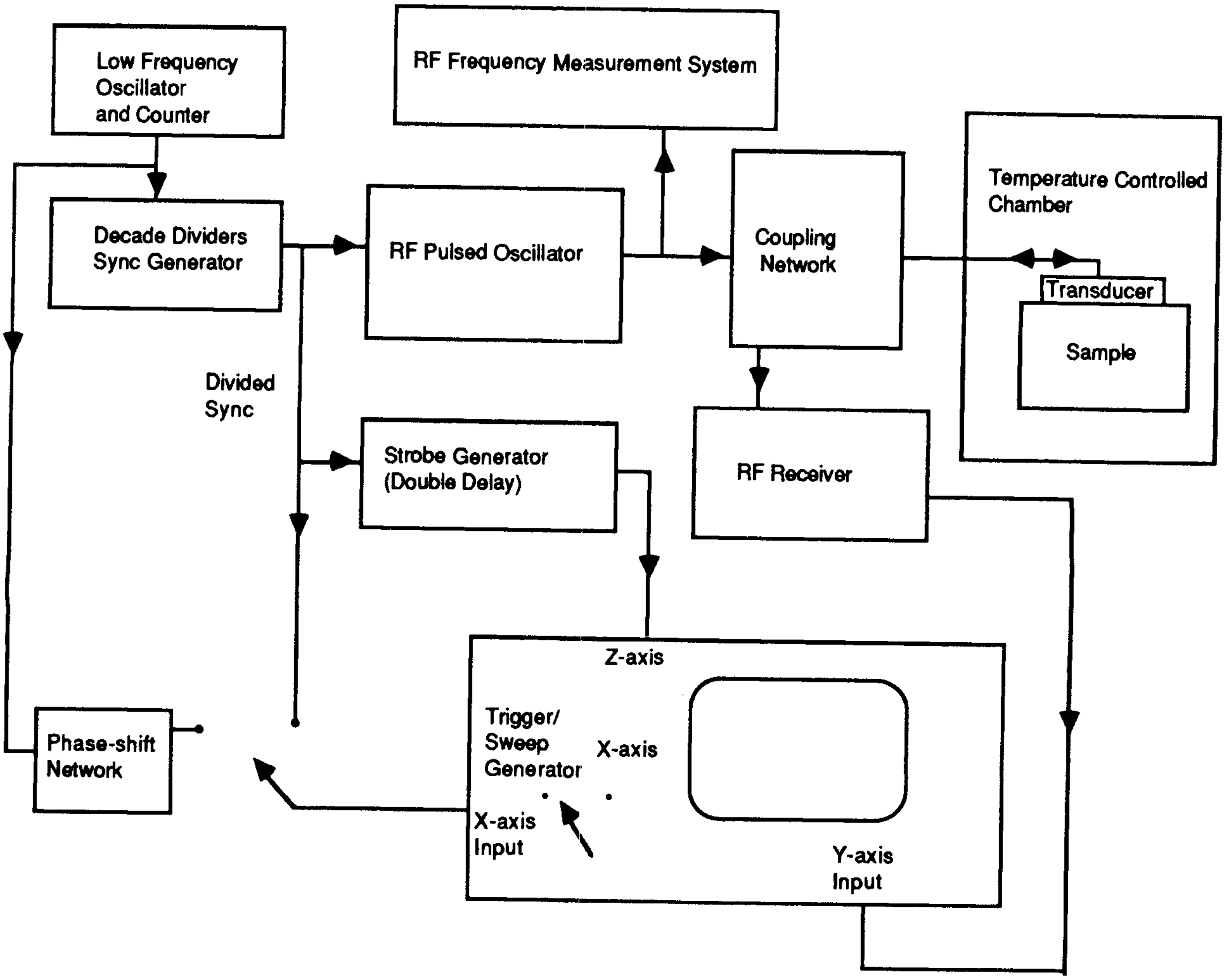

Low Frequency Oscillator and Counter
RF Frequency Measurement System
Decade Dividers Sync Generator
RF Pulsed Oscillator
Coupling Network
Temperature Controlled Chamber
Transducer
Sample
Divided Sync
Strobe Generator (Double Delay)
RF Receiver
Phase-shift Network
Z-axis
Trigger/ Sweep Generator
X-axis
X-axis Input
Y-axis Input

Fig. 3 Equipment arrangement for echo–overlap method.[9,20,21]

reference oscillator. The received pulse (an echo in a single-transducer arrangement) is combined with the CW reference signal in a phase-sensitive detector. If the relative phase between the received acoustic signal and the CW reference signal is then changed, the output of the phase-sensitive detector varies in response. The change in relative phase may result, for example, from a variation in the acoustic path or from a variation in the sound velocity as a function of temperature.

Two basic means are reported to measure the change in relative phase. One means employs a phase shift network that monitors the amount of phase added to the reference signal necessary to maintain a constant output from the phase-sensitive detector.[22,23] The frequency of the CW reference signal is held fixed during this process. The second means uses a change in the frequency of the CW reference signal to maintain a constant detector output.[24,25] In the pulse phase-locked loop system a voltage-controlled oscillator network replaces the phase shift network.[25]

In both the constant-frequency and variable-frequency techniques, the total phase seen by the phase detector includes acoustic contributions related to the acoustic wavelength in sample length as well as collective nonacoustic contributions. The frequency-dependent nonacoustic contributions include those from electrical signal propagation paths and phase perturbations associated with various electrical components. In measurements of phase velocity in fluids, for example, the change in phase corresponding to a given change in path length may be measured. For the constant-frequency technique both the wavelength and the nonacoustic phase term are constants. The phase velocity is obtained from a measurement of the slope of the linear curve giving the change in phase as a function of the change in path length. For the variable-frequency technique the wavelength and the nonacoustic phase term are not fixed and the velocity can be calculated only after correcting for the variations in the nonacoustic phase term.

Variations in the pulse transit time can be measured with a sensitivity of parts in 10^8 with pulse interferometer methods. Absolute velocity measurements can be made to accuracies that are comparable to the pulse superposition and echo-overlap methods.

3 CONTINUOUS-WAVE METHODS

Continuous-wave methods are among the most expedient for measuring sound velocity and attenuation of materials with appropriate geometry. Consider a sample of length $l/2$ with flat and parallel faces and assume that one face is driven continuously by a noncontacting transducer with particle velocity $v = v_0 \cos \omega t$. The multiple reflections of the wave from the parallel sample faces results in a steady-state particle velocity amplitude at the position of the driving transducer given by[6,26]

$$v = \frac{v_0 e^{\alpha l/2}}{\sqrt{2}(\cosh \alpha l - \cos kl)^{1/2}}, \tag{2}$$

where α is the attenuation coefficient in nepers per centimetres and k is the propagation constant. The amplitude response thus consists of a series of mechanical resonances (standing waves) whose frequencies correspond to an integral number of half wavelengths within the sample length. The condition for the nth resonance is $kl = 2\pi n(\alpha l \ll 1)$, where n is an integer.

For materials having a fixed length and sufficiently small dispersion the phase velocity c_p can be determined from the frequency difference $f_n - f_{n-1}$ between two adjacent resonances as $c_p = l(f_n - f_{n-1})$. The attenuation α can be calculated from a determination of the frequency width at half power, $\Delta\omega$, as $\alpha = \Delta\omega/2c_p$. These results are based on the assumption that the sample responds as a simple resonator driven by noncontacting transducers. If contacting transducers are used, the sample and transducers form a composite resonating system whose resonant frequencies are predicted by the roots of a transcendental equation. Approximate solutions to the transcendental equation have resulted in correction factors for the above formulas.[27] However, errors in velocity determination can exceed 10% under certain conditions.

In the method of ultrasonic resonance spectroscopy[28,29] measurements of the modal frequencies for a sample of given geometry are compared to predictive analytical models of the sample modal frequencies. Analysis of the amplitude–frequency spectrum allows an assessment of the elastic constants as well as the internal friction (attenuation) of the sample.

For gases, fixed-frequency CW methods such as the resonating-tube technique and Kundt's dust-tube method are useful for velocity measurements. In the latter method cork dust initially distributed uniformly through the tube piles up at the nodes of a standing wave set up in the tube. The distance between the nodes are used to determine the wavelength λ in a gaseous medium, and the sound velocity c_p is calculated from $c_p = f\lambda$. In the resonating-tube method, a tuning fork or other single-frequency source of sound held at the open end of a gaseous column produces sound reinforcement whenever the vibrational mode frequency of the column is equal to the vibrational frequency of the source. If the column length is then varied by a half wavelength, sound reinforcement again occurs. A high-frequency transducer mounted on a micrometer screw can yield an accurate determination of the speed of sound in relatively small amounts of gas.

A particularly useful aspect of CW methods is the determination of small changes in the attenuation or sound velocity. Specific frequencies can be chosen to isolate small changes in the propagation constant or attenuation resulting from a change in some intrinsic material variable.[26]

4 OPTICAL METHODS

Optical techniques can be used for sound velocity measurements in transparent media. The methods generally are based on the optical determination of acoustic wavelength either from diffraction or scattering of light by the sound waves. The wavelength together with the acoustic frequency allows determination of phase velocity. For techniques based on diffraction it is essential to know the type of diffraction experienced by the light.

4.1 Diffraction of Light by Sound (Fraunhofer Diffraction Conditions for Light)

Consider an ultrasonic transducer generating compressional waves in a transparent propagation medium. If the propagating sound wave encounters light traversing the same medium at some angle incident to the sound wave, the light is diffracted. Assuming that the Fraunhofer conditions hold for the light diffraction (i.e., parallel light rays are brought to a focus), two distinct physical processes can produce the diffraction effects. One involves the formation of a corrugation in the phase fronts of the light due to the acoustically induced spatial variation of the index of refraction. This phenomenon is called Raman–Nath diffraction.[30,31] The second process involves the reflection of light from the (evenly spaced) crests of the sound wave that occurs under conditions similar to that of x-ray diffraction. This phenomenon is called Bragg diffraction.[32] Both types of diffraction can be used to determine the sound velocity of the material.

A dimensionless parameter has been introduced to determine which type of diffraction predominates for given experimental conditions. The parameter Q is defined by[33]

$$Q = \frac{k^2 L}{\mu_0 k^*}, \tag{3}$$

where k is the ultrasonic propagation constant, L is the width of the ultrasonic beam, μ_0 is the index of refraction, and k^* is the propagation constant of the light in vacuo. If $Q > 9$, Bragg diffraction is predominate. If $Q < 1$, Raman–Nath diffraction occurs. If $1 < Q < 9$, the diffraction is mixed. A typical transparent solid has an acoustic phase velocity of the order 5×10^3 m/s and an index of refraction of order 1.5. Consider an ultrasonic beam 1.27×10^{-2} m in width and illuminated with light of wavelength 632.8 nm. If the ultrasonic frequency is approximately 27 MHz or less, the interaction satisfies Raman–Nath diffraction conditions, according to Eq. (3). If the frequency is greater than approximately 81 MHz, the interaction is governed by Bragg diffraction. In liquids the demarcation frequencies are often quite different from solids. In water for the same ultrasonic beam width and wavelength of light as the above example, Raman–Nath diffraction occurs for acoustic frequencies less than 7.7 MHz while Bragg diffraction occurs above 23 MHz.

Raman–Nath Diffraction For Raman–Nath diffraction with the light incident at an angle $\pi/2 - \phi$ to the sound beam direction, the light is diffracted into diffraction orders given by the expression[6]

$$\sin(\theta_n + \phi) - \sin(\phi) = \frac{n\lambda^*}{\lambda}, \tag{4}$$

where θ is the angle of diffraction with respect to the incident beam, λ^* is the wavelength of light in the medium, λ is the ultrasonic wavelength, and n is an integer denoting the diffraction order. The ultrasonic wavelength can be determined from Eq. (4).

Bragg Diffraction For Bragg diffraction, the angle of diffraction θ_n is set equal to twice the negative angle of incidence $(-\phi)$ such that

$$n\lambda^* = 2\lambda \sin \phi_B, \tag{5}$$

where ϕ_B is the Bragg angle. Bragg diffraction is used to measure wave velocities in the frequency range from approximately 100 MHz to several gigahertz. The technique has been used to measure wave velocity, including local velocities,[34] to an accuracy estimated to be better than 0.1%. Measurements with precision better than 0.01% in homogeneous samples have been reported.

4.2 Direct Measurements (Fresnel Diffraction Conditions for Light)

Consider a standing ultrasonic wave in a sample through which collimated light perpendicular to the sound beam is passed. A measuring microscope or similar optical device is focused so that the image of the sound wave can be viewed at the instrument's focal plane. The light that is diffracted by the sound wave is viewed under experimental conditions such that nonparallel light rays are focused to form the image (Fresnel diffraction). Because the images of the wavefronts are λ/2 apart under such conditions, it is possible to determine directly the wavelength in the medium. The technique[35] is sufficiently sensitive to detect local variations in phase velocity greater than 0.01%.

4.3 Scattering Techniques: Brillouin Scattering

Vibrational Modes in a Crystal Lattice Lattice vibrations are caused by thermal and other excitations of the lattice. These vibrations form vibrational modes called normal modes within the lattice and play a major role in determining the thermal properties (e.g., heat capacity) of the solid. The normal modes form two major branches. The acoustic branch includes normal-mode frequencies that vanish linearly as the propagation vector k approaches zero (long-wavelength limit). The optical branch includes normal-mode frequencies that do not vanish as the propagation vector approaches zero.

The quantum mechanical treatment of the lattice vibrations predicts that the normal modes gain and lose energy in quantized units just as in the case of the harmonic oscillator. Because of the similarity of the lattice vibrations to the electromagnetic radiation field in a cavity, a quantum of lattice vibrational energy called the phonon is defined, primarily for convenience, in analogy to the electromagnetic quantum of energy called the photon. As with photons, phonons are given the property that each phonon carries an energy $\hbar\omega$, where $\hbar$ is Planck's constant divided by 2π, and under certain conditions[36] behaves as if it carries a crystal momentum of value $\hbar k$. The reader is referred to Chapter 57 for detail.

Interaction of Light with Vibrational Modes Light interacting with the normal modes in the material can gain or lose energy and momentum. Brillouin scattering occurs when the modes of the acoustic branch interact with light. Consider the interaction of a photon with a material whose index of refraction is μ_0. Application of the laws of conservation of energy and momentum, respectively, to the scattering process gives

$$\hbar\omega^{*\prime} = \hbar\omega^{*} + \hbar\omega, \qquad \hbar\mu_0 k^{*\prime} = \hbar\mu_0 k^{*} + \hbar k + \hbar K, \quad (6)$$

where $\omega^{*\prime}$ and $k^{*\prime}$ are the frequency and propagation vector, respectively, of the incident photon; ω^{*} and k^{*} are the frequency and propagation vector of the scattered photon; ω and k are the frequency and crystal momentum assigned to the phonon; K is the reciprocal lattice vector (assumed to be zero in this case); and μ_0 is the index of refraction of the material. Since the photon frequency shifts are small for this process, the propagation vectors are very nearly the same length. The vector addition expressed by Eq. (6) gives an isosceles triangle whose apex angle is θ, the complement of the scattering angle. Equation (6) leads to the result[37]

$$\omega = \pm 2\omega^{*}\left(\frac{c}{C_l}\right)\sin\left(\frac{\theta}{2}\right), \quad (7)$$

where C_l is the speed of the photon in the medium, c is the speed of the phonon, and θ is the scattering angle. The frequency ω is equal to the difference in frequencies between the incident and scattered photons.

The technique produces a scattered photon spectra of three lines, one of which is unshifted. The two frequency-shifted lines are called the Stokes line (a phonon is created) and the anti-Stokes line (a phonon is annihilated). These lines can be separated and measured with optical devices such as the Fabry–Perot interferometer to obtain ω. If desirable, the Bragg condition can be used for constructive reinforcement[38] by adjusting θ. Equation (7) can be solved for the velocity c in terms of the other quantities measured.

Brillouin scattering is useful in the investigation of the sound velocity of solids and liquids, especially near a phase transition where attenuation can become very large.[39] The technique is also advantageous in that it is one of the few techniques capable of assessing sound velocity at frequencies of several gigahertz. The technique has been used for sound velocity measurements for both longitudinal and mixed (longitudinal–transverse) modes. Uncertainties in the index of refraction, the scattering angle, and line width of the Stokes and anti-Stokes lines influence the accuracy of the measurement. The precision is considerably better, however, with values of 0.1% possible.

5 SOURCES OF ERROR

A number of sources of error arise in the measurements of velocity and attenuation. Their effect on the measurements cannot be overstated. Errors arising from time or length measurements can be evaluated directly, but among the less easily evaluated errors are those arising from velocity dispersion, material inhomogeneity, phase cancellation resulting from lack of parallelism of the transducers, diffraction, and the effects of thickness of the transducer and transducer bond. Many aspects of the sources of error are best addressed in the context of the measurement technique used, but certain characteristics of the error sources have more generic implications that are considered here.

5.1 Dispersion

In a linear, dispersive, propagation medium the acoustic waveform $u(x, t)$ at position x and at time t in the medium is given by the expression[40]

$$u(x, t) = F(x, t)e^{i\Omega(x,t)} + \text{c.c.}, \qquad \Omega(x, t) = kx - \omega t, \quad (8)$$

providing

$$\frac{d\omega}{dk} - \frac{x}{t} = 0, \qquad \frac{d^2\omega}{dk^2} \neq 0, \quad (9)$$

where ω is the angular frequency and c.c. is the complex conjugate. The functional dependence of the angular frequency on the wavenumber k defines the dispersion relation. The condition $d^2\omega/dk^2 \neq 0$ is the defining condition for a dispersive propagation medium and $d\omega/dk = c_g$ is defined to be the group velocity. For a given choice of x and t, Eq. (9) may be solved for k if the dispersion function $\omega(k)$ is known. Thus, different parts of the waveform corresponding to different given pairs (x, t) propagate with different group velocities $c_g(k)$ corresponding to specific wavenumbers determined from Eq. (9).

Both the complex amplitude $F(x, t)$ and the phase function $\Omega(x, t)$ in Eq. (8) are functionally dependent on x and t. Equation (8) has the form of an elementary waveform, but $F(x, t)$ is actually a position- and time-varying envelope of the waveform with the position- and time-varying phase $\Omega(x, t)$ defining nonuniformly spaced maxima and minima of the wave. Equation (8) thus defines a disturbance that in general changes shape during its propagation through a dispersive medium.

The propagation of a given constant value of phase Ω_0 defined by $\Omega(x, t) = \Omega_0$ moves with the phase velocity $c_p(x, t) = dx/dt = \omega/k$. The phase velocity $c_p(x, t)$ is generally different for different values of Ω_0 and for different values of x and t. An observer following a particular phase (e.g., crest of a wave) moves with the local phase velocity having a locally varying frequency and wavenumber. Thus, the phase velocity of a wave of finite length propagating in a dispersive medium can only be defined locally (i.e., for given x and t) and for a given phase Ω_0. Hence, if a given pulse measurement technique follows a given phase Ω_0 of the waveform at different monotonically increasing or decreasing values of the space–time coordinates (x_i, t_i), then the local phase velocity at (x_α, t_α) for phase Ω_0 can be obtained by plotting x_i as a function of t_i and measuring the tangent to the curve at (x_α, t_α). For dispersive media the curve is generally nonlinear, but for nondispersive media the curve is linear and coincides with the curve for the group velocity $c_g = x/t$. Techniques employing a change in the operation frequency of the system cannot generally be used to accurately measure phase velocity in highly dispersive media, since the shift in operation frequency violates conditions under which the phase velocity is defined.

Measurements using gated continuous waves (i.e., tonebursts) of finite length can be made to approximate CW conditions if the pulse lengths are sufficiently long. The appropriate pulse length L is dependent on the intrinsic error M of the measurement technique and is given by the expression[22]

$$L \geq \frac{\pi(d^2\omega/dk^2)}{Mc_p}. \quad (10)$$

5.2 Diffraction

Because the transducer is of finite size, the acoustic wave generated in the sample spreads out into a diffraction field that can introduce error in both attenuation and velocity measurements. The effects of diffraction are usually treated[41–44] for the case of circular, axially concentric transmitting and receiving transducers of the same radius a. The transmitting transducer is treated as a finite piston source in a semi-infinite medium. The acoustic field is found at each point in the propagation medium, and an integration is performed over the area in the field presented by the receiving transducer. In one treatment the effects of diffraction can be represented for waves of wavelength λ propagating along the z-direction in an isotropic medium by the expression[43]

$$u(z, t) = 2(I_x^2 + I_y^2)^{1/2}\cos(kc_pt - kz - \phi),$$
$$\phi = \tan(I_y/I_x), \quad (11)$$

where the quantities I_x and I_y are integral functions of the dimensionless parameter $S = z\lambda/a^2$.

These equations show that the diffraction error involves an intensity variation in the diffraction pattern and a distortion of the otherwise planar wavefront. Phase variations in the surface plane of the receiving transducer give rise to interference effects and a loss in signal when integrated over a phase-sensitive transducer surface. The factor ϕ in Eq. (11) produces a variation in the phase of the propagating wave that results in an error in the velocity measurements, if not properly corrected.

5.3 Phase Cancellation

A variation in the phase of the wave over the area presented by the surface of a phase-sensitive transducer of radius *a* can give rise to phase cancellation of different portions of the received signal and set limits on the measurement accuracy of both velocity and attenuation. Sources of such phase variation include diffraction, media inhomogeneity, temperature gradients, wedging, and nonparallelism of the sample surfaces. Nonparallelism of opposite surfaces of the sample by an angle θ is representative of the effects of phase cancellation. For example, in measurements employing acoustic pulses of wavenumber k reflecting from the ends of the sample the amplitude of the pulse received after n double reflections is diminished by the factor $J_1(2kan\theta)/kan\theta$ due to nonparallelism.[5] The use of phase-insensitive devices such as the "acoustoelectric transducer" can minimize phase cancellation error,[45] although such devices are generally less sensitive than the commonly used phase-sensitive piezoelectric elements.

5.4 Transducer Bond

Major sources of error in measurements of velocity are the effects of the transducer and the bond coupling the transducer to the sample. Both the transducer and bond introduce additional phase into the measurement system, which can lead to errors of the order of parts in 10^2 in the measurement of velocity. Methods for correcting such errors are well established and have been addressed as appropriate to the specific measurement techniques discussed. An alternative to such corrections is the use of noncontacting transducers such as capacitive[46] and electromagnetic transducers[47] and optical techniques.

6 CONCLUSION

The methods presented in this chapter for measuring velocity and attenuation in solids, liquids, and gases are basic and instructive. It is not possible to single out a best method to use in all measurement situations. Considerations of sample dimensions, environmental constraints, and specific physical and chemical properties of the material play a role in selecting the measurement method. Many variations of the basic methods delineated in this chapter have been developed, some of which are useful in specific applications.[1–6]

It must be emphasized that any method is only as good as the care taken by the investigator to correct for sources of error. For example, a critical analysis of the accuracies of the pulse superposition and the echo-overlap methods shows that absolute accuracies of the order of parts in 10^6 are possible for velocity measurements with either method if care is taken to correct for diffraction, wedging, transducer and bond thicknesses, and other sources of error. Similar accuracies are expected to be achievable with other methods such as pulse interferometry and some of the CW methods.

In addition to the error sources discussed here, a serious limitation on the accuracy of absolute velocity measurements is the determination of the length of the sample. Length considerations are not a factor in relative velocity measurements where sensitivities of parts in 10^8 are theoretically possible. For attenuation measurements error corrections are no less critical. The aforementioned error sources and the intrinsic limitations on instrument calibration for measuring wave amplitudes geneally limit the accuracy of attenuation measurements to parts in 10^3 even after corrections are applied.

Careful sample preparation and apparatus set-up together with the appropriate error corrections and efforts to control equipment instability and environmental factors (such as temperature and pressure) allow measurements of velocity, relative velocity, and attenuation that approach the theoretical limits for a given method.

REFERENCES

1. H. J. McSkimin, in W. P. Mason (Ed.), *Physical Acoustics*, Vol. IA, Academic, New York, 1964, pp. 271–417.
2. E. P. Papadakis, in W. P. Mason (Ed.), *Physical Acoustics*, Vol. IVB, Academic, New York, 1968, pp. 269–328.
3. E. P. Papadakis, in W. P. Mason and R. N. Thurston (Eds.), *Physical Acoustics*, Vol. XI, Academic, New York, 1975, pp. 151–211.
4. E. P. Papadakis, in W. P. Mason and R. N. Thurston (Eds.), *Physical Acoustics*, Vol. XII, Academic, New York, 1976, pp. 277–374.
5. R. Truell, C. Elbaum, and B. B. Chick, *Ultrasonic Methods in Solid State Physics*, Academic, New York, 1969.
6. M. A. Breazeale, J. H. Cantrell, Jr. and J. S. Heyman, in P. H. Edmonds (Ed.), *Methods of Experimental Physics, Ultrasonics*, Vol. 19, Academic, New York, 1981, pp. 67–135.
7. F. A. Firestone and J. R. Frederick, "Refinements in Supersonic Reflectoscopy. Polarized Sound," *J. Acoust. Soc. Am.*, Vol. 18, 1946, pp. 200–211.
8. J. R. Pellam and J. K. Galt, "Ultrasonic Propagation in Liquids: I. Application of Pulse Technique to Velocity and Absorption Measurements at 15 Megacycles," *J. Chem. Phys.*, Vol. 14, 1946, pp. 608–614.
9. E. P. Papadakis, "Ultrasonic Phase Velocity by the Pulse-Echo-Overlap Method Incorporating Diffraction Phase Corrections," *J. Acoust. Soc. Am.*, Vol. 42, 1967, pp. 1045–1051.
10. H. J. McSkimin, "Measurement of Ultrasonic Wave

Velocity for Solids in the Frequency Range 100–500 Mc," *J. Acoust. Soc. Am.*, Vol. 34, 1962, pp. 404–409.

11. R. L. Roderick and R. Truell, "The Measurement of Ultrasonic Attenuation in Solids by the Pulse Technique and Some Results in Steel," *J. Appl. Phys.*, Vol. 23, 1952, pp. 267–279.
12. J. H. Andreae, R. Bass, E. L. Heasell, and J. Lamb, "Pulse Techniques for Measuring Ultrasonic Absorption in Liquids," *Acustica*, Vol. 8, 1958, pp. 131–142.
13. B. Chick, G. Anderson, and R. Truell, "Ultrasonic Attenuation Unit and Its Use in Measuring Attenuation in Alkali Halides," *J. Acoust. Soc. Am.*, Vol. 32, 1960, pp. 1186–1193.
14. N. P. Cedrone and D. R. Curran, "Electronic Pulse Method for Measuring the Velocity of Sound in Liquids and Solids," *J. Acoust. Soc. Am.*, Vol. 26, 1954, pp. 963–966.
15. R. L. Forgacs, "An Improved Sing-Around System for Ultrasonic Velocity Measurements," *IRE Trans. Instrum.*, Vol. 9, 1960, pp. 359–367.
16. J. Williams and J. Lamb, "On the Measurement of Ultrasonic Velocity in Solids," *J. Acoust. Soc. Am.*, Vol. 30, 1958, pp. 308–313.
17. M. Redwood and J. Lamb, "On the Measurement of Attenuation in Ultrasonic Delay Lines," *Proc. Inst. Electr. Eng.*, Vol. 103B, 1956, pp. 773–780.
18. H. J. McSkimin, "Variations of the Ultrasonic Pulse-Superposition Method for Increasing the Sensitivity of Delay-Time Measurements," *J. Acoust. Soc. Am.*, Vol. 37, 1965, pp. 864–871.
19. H. J. McSkimin and P. Andreatch, Jr., "Measurement of Very Small Changes in the Velocity of Ultrasonic Waves in Solids," *J. Acoust. Soc. Am.*, Vol. 41, 1967, pp. 1052–1057.
20. J. E. May, Jr., "Precise Measurement of Time Delay," *IRE Nat. Conv. Rec.*, Vol. 6, 1958, pp. 134–142.
21. E. P. Papadakis, "Ultrasonic Attenuation and Velocity in Three Transformation Products in Steel," *J. Appl. Phys.*, Vol. 35, 1964, pp. 1474–1482.
22. W. T. Yost, J. H. Cantrell, and P. Kushnick, "Fundamental Aspects of Pulse Phase-Locked Loop Technology-Based Methods for Measurement of Ultrasonic Velocity," *J. Acoust. Soc. Am.*, Vol. 91, 1992, pp. 1456–1468.
23. W. M. Whitney and C. E. Chase, "Velocity of Sound in Liquid Helium at Low Temperatures," *Phys. Rev. Lett.*, Vol. 9, 1962, pp. 243–245.
24. R. J. Blume, "Instrument for Continuous High Resolution Measurement of Changes in the Velocity of Ultrasound," *Rev. Sci. Instrum.*, Vol. 34, 1963, pp. 1400–1407.
25. J. S. Heyman and E. J. Chern, "Ultrasonic Measurement of Axial Stress," *J. Testing Evaluation*, Vol. 10, 1982, pp. 202–211.
26. D. I. Bolef and J. G. Miller, in W. P. Mason and R. N. Thurston (Eds.), *Physical Acoustics*, Vol. VIII, Academic, New York, 1971, pp. 96–201.
27. E. J. Chern, J. H. Cantrell, Jr., and J. S. Heyman, "Improved Formula for Continuous-Wave Measurements of Ultrasonic Phase Velocity," *J. Appl. Phys.*, Vol. 52, 1981, pp. 3200–3204.
28. O. Anderson, "Rectangular Parallelepiped Resonance-A Technique of Resonance Ultrasound and Its Applications to the Determination of Elasticity at High Temperatures," *J. Acoust. Soc. Am.*, Vol. 91, 1992, pp. 2245–2253.
29. H. Ledbetter, C. Fortunko, and P. Heyliger, "Elastic Constants and Internal Friction of Polycrystalline Copper," *J. Mater. Res.*, Vol. 10, 1995, pp. 1352–1353.
30. C. V. Raman and N. S. N. Nath, "The Diffraction of Light by High Frequency Sound Waves: Part 1," *Proc. Indian Acad. Sci.*, Sect. A, Parts 1 and 2, Vol. 2, 1935, pp. 406–420.
31. C. V. Raman and N. S. N. Nath, "The Diffraction of Light by High Frequency Sound Waves: Part V. General Considerations—Oblique Incidence and Amplitude Changes," *Proc. Indian Acad. Sci.*, Sect. A, Part V, Vol. 3, 1936, pp. 459–465.
32. A. B. Bhatia and W. J. Noble, "Diffraction of Light by Ultrasonic Waves I. General Theory," *Proc. R. Soc. London*, Ser. A 220, 1953, pp. 356–368.
33. W. R. Klein, B. D. Cook, and W. G. Mayer, "Light Diffraction by Ultrasonic Gratings," *Acustica*, Vol. 15, 1965, pp. 67–74.
34. F. Michard and B. Perrin, "New Optical Probe for Local Elastic Study in Transport Media," *J. Acoust. Soc. Am.*, Vol. 64, 1978, pp. 1447–1456.
35. W. G. Mayer and E. A. Hiedemann, "Optical Methods for the Ultrasonic Determination of the Elastic Constants of Sapphire," *J. Acoust. Soc. Am.*, Vol. 30, 1958, pp. 756–760.
36. C. Kittel, *Introduction to Solid State Physics*, 3rd ed., Wiley, New York, 1966, pp. 134ff.
37. N. W. Ashcroft and N. D. Mermin, *Solid State Physics*, Saunders, Philadelphia, 1976, pp. 480–483.
38. H. F. Pollard, *Sound Waves in Solids*, Pion Limited, London, 1977, pp. 308–309.
39. P. A. Fleury, in W. P. Mason and R. N. Thurston (Eds.), *Physical Acoustics*, Vol. VI, Academic, New York, 1970, pp. 37–58.
40. G. B. Whitham, *Linear and Nonlinear Waves*, Wiley-Interscience, New York, 1974.
41. H. Seki, A. Granato, and R. Truell, "Diffraction Effects in the Ultrasonic Field of a Piston Source and Their Importance in the Accurate Measurement of Attenuation," *J. Acoust. Soc. Am.*, Vol. 28, 1956, pp. 230–238.
42. A. S. Khimunin, "Numerical Calculation of the Diffraction Corrections for the Precise Measurement of Ultrasound Absorption," *Acustica*, Vol. 27, 1972, pp. 173–181.
43. G. C. Benson and O. Kiyohara, "Tabulation of Some Integral Functions Describing Diffraction Effects in the Ultrasonic Field of a Circular Piston Source," *J. Acoust. Soc. Am.*, Vol. 55, 1974, pp. 184–185.

44. E. P. Papadakis, "Ultrasonic Diffraction Loss and Phase Change in Anisotropic Materials," *J. Acoust. Soc. Am.*, Vol. 40, 1966, pp. 863–876.

45. J. S. Heyman, "Phase Insensitive Acoustoelectric Transducer," *J. Acoust. Soc. Am.*, Vol. 64, 1978, pp. 243–249.

46. J. H. Cantrell, Jr., and M. A. Breazeale, "Elimination of Transducer Bond Corrections in Accurate Ultrasonic Wave Velocity Measurements by Use of Capacitive Transducers," *J. Acoust. Soc. Am.*, Vol. 61, 1977, pp. 403–406.

47. D. J. Meridith, R. J. Watts-Tobin, and E. R. Dobbs, "Electromagnetic Generation of Ultrasonic Waves in Metals," *J. Acoust. Soc. Am.*, Vol. 45, 1969, pp. 1393–1401.

56

ULTRASONIC RELAXATION PROCESSES

HENRY E. BASS AND F. DOUGLAS SHIELDS

1 INTRODUCTION

Studies of acoustic absorption and dispersion have proven to be an effective tool in developing a physical description of molecular interactions in gases and liquids. Experimentally, absorption or velocity dispersion is measured as a function of frequency (or, in gases, as a function of frequency and pressure). These same quantities are computed using the theory outlined below, which requires some insight into the particular physical processes that give rise to the absorption or dispersion (e.g., vibrational relaxation). Computed values are compared to experiment, and the assumed physical process or interaction parameters are varied until agreement is achieved. The theoretical model that gives rise to predicted absorption and dispersion that agree with experiment are assumed to correctly represent the interactions taking placed at the microscopic level.

This chapter is concerned with ultrasonic relaxation processes. The relaxation contribution to absorption of sound in air at lower frequencies is addressed in Chapter 32 of Volume 1. One source of relaxation absorption is that associated with chemical reactions, especially in fluids. That type of relaxation is not considered here. This chapter will be concerned, primarily, with the theory that relates the microscopic excitation/deexcitation interactions to acoustic absorption or dispersion.

2 PHENOMENOLOGICAL EXPLANATION OF RELAXATION

For the purpose of developing an understanding of relaxation processes, first consider an ideal gas made up of diatomic molecules. The individual molecules are free to move translationally in three directions, rotate about two perpendicular axes (actually three but the third has zero moment of inertia and so has no energy), and vibrate along the bond joining the atoms. Some energy is associated with each of these allowed motions.

Translational motion can be considered nonquantized. Any energy is allowed. As the molecules translate, they collide, exchanging energy with their collision partners. At atmospheric pressure molecular collisions take place at a rate of about 10^{11} s^{-1}. A single collision is typically sufficient to transfer translational energy from one molecule to another. However, a certain period of time is required to randomize energy associated with excess velocity in a particular direction. This time is often referred to as the translational relaxation time. Kohler, following Maxwell,[1] associates viscosity with this relaxation time writing $\tau_{\text{tr}} = \eta/p = 1.25\tau_c$, where p is gas pressure and τ_c is the time between collisions.

As the pressure is lowered, the rate of collisions decreases proportionately. At 1 Torr (1/760 atm), the translational relaxation time is about 10^{-8} s; at 1 mTorr it is 10^{-5} s. These time regimes can be effectively studied using ultrasonics.

Unlike translational motion, rotation and vibration are noticeably quantized. During a collision, a change in rotational or vibrational state can only occur if the change in energy of another state is sufficient to allow at least one quantum jump. For rotational energy transfer, the spacing between energy levels is given by $2(J+1)B$, where J is the rotational quantum number $B = \hbar^2/2I$ and I is the effective moment of inertia. If one assumes J is the most probable value (from a Boltzman distribution), the value of $2(J+1)B$ for a typical molecule (say N_2) in temperature units is about 1°. This means that in a gas above 1 K, essentially all collisions will have sufficient translational energy to cause multiple changes in joules.

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker
Copyright © 1997 John Wiley & Sons, Inc.

As a result, rotation rapidly equilibrates with translation.

An exception is hydrogen, which has much larger rotational energy level spacing due to the small moment of inertia. On the average, as many as 350 collisions[2] may be necessary to transfer a quantum of rotational energy in H_2. At a pressure of 1 atm, this gives a relaxation time of about 2×10^{-8} s. It should be noted that a given collision does or does not transfer a quantum of energy. The 350-collision average means that only 1 collision in 350 has the proper geometry and energy to cause a transfer of one quantum of rotational energy. The number of collisions necessary, on the average, to transfer one quantum of energy is referred to as the collision number Z. When rotational energy is involved, the subscript "rot" is typically added (Z_{rot}). The inverse of this dimensionless quantity is the probability of transferring a quantum in a collision (P_{rot}). Since rotational energy level spacings are unequal, a $1 \rightarrow 2$ transition should be more probable than a $2 \rightarrow 3$ transition. These events are distinguished by using the symbols $P_{\text{rot}}^{1\rightarrow 2}$ or $P_{\text{rot}}^{2\rightarrow 3}$.

Generally speaking, the probability for transferring a quantum of energy in a collision decreases rapidly with the size of the quantum transferred. Since vibrational levels are much more widely spaced than rotational energy levels, vibrational relaxation times are much longer than rotational. Vibrational levels in a single vibrational mode are approximately equally spaced. This means that energy can be exchanged between levels (i.e., the vibrational quantum number goes up in one molecule and down in the other) with very little energy exchanged between vibration and translation. The result is that such vibration-to-vibration exchanges take place very rapidly. The vibrational relaxation time is controlled by the time it takes energy to transfer between translation and the lowest lying vibrational level. Since this energy level varies greatly for different molecules, so do the probabilities of vibrational energy transfer during a collision. During N_2 collisions with N_2, Z_{10} is near 1.5×10^{11}, so the relaxation time is near 15 s[3] (Z_{10} is the number of collisions needed to transfer energy from the lowest vibrational level to translation). Large molecules have vibrational energy levels that are very close together. A molecule such a C_2H_6 requires only 100 collisions to transfer a quantum of vibrational energy from the first vibration level into translation.[4] This very wide range of relaxation times presents interesting experimental challenges.

To this point, there has been no attempt to rigorously define relaxation times in terms of energy transfer probabilities. In fact, such a relation is possible in a simple form for only the few cases where gases exhibit a single relaxation time. Nevertheless, the simple case provides valuable insight into the behavior of more complex systems and deserves a detailed description.

Consider the case where the population of a vibrational state is excited to an energy E_v that is greater than the energy $E_v(T_{\text{tr}})$, which it could have in Boltzmann equilibrium with translation. In this case, the excess vibrational energy will equilibrate with translation according to a standard relaxation equation,

$$-\frac{dE_v}{dt} = \frac{1}{\tau}\,[E_v - E_v(T_{\text{tr}})]. \qquad (1)$$

The return to equilibrium occurs due to energy transfer during individual collisions.

The rate of energy transfer k_{10} is defined as the rate at which molecules go from the first excited state to the ground state due to collisions at a pressure of 1 atm. This rate is just the collision frequency M times the probability of energy transfer, $P^{1\rightarrow 0}$, times the mole fraction of molecules in the first excited state, x_1. During some collisions, the reverse process will occur, that is, some molecules in the ground state will become excited at a rate k_{01}. In equilibrium, equal numbers of molecules go in both directions, so

$$k_{10}x_1 - k_{01}x_0 = 0. \qquad (2)$$

As explained above, energy is quickly shared from the first excited level of the vibrational mode to higher levels of the mode by vibration-to-vibration exchanges. Assuming quantum mechanical laws hold for probabilities of energy exchanges between vibrational levels of a harmonic oscillator, Landau and Teller showed[5] that

$$-\frac{dE_v}{dt} = k_{10}(1 - e^{-h\upsilon/kT})[E_v - E_v(T_{\text{tr}})], \qquad (3)$$

where υ is the vibrational frequency of the relaxing mode. By comparison with Eq. (1),

$$\tau = \frac{1}{k_{10}(1 - e^{-h\upsilon/kT})}.$$

The link between the relaxation time and ultrasonic absorption and dispersion is understood by noting that the relaxation process makes the specific heat of the gas time (or frequency) dependent. This time dependence can be obtained from the energy relaxation equation. Consider, as above, that the specific heat of a simple gas can be divided into translational, rotational, and vibrational contributions. For now, assume that the translation and rotational energy both equilibrate rapidly enough to follow any acoustically induced temperature variations. In this case, the effective specific heat can be written as

$$(C_v)_{\text{eff}} = C_v^\infty + C' \frac{dT'}{dT_{\text{tr}}}, \tag{4}$$

where C_v^∞ is the sum of rotational and translational specific heats, C' is the relaxing specific heat, and T' is the instantaneous temperature of the relaxing mode (in this case, vibration). From the energy relaxation equation [Eq. (3)], for small periodic variations in T_{tr} and T_v about their equilibrium values,

$$(C_v)_{\text{eff}} = C_v^\infty + \frac{C'}{1 + i\omega\tau}, \tag{5}$$

where ω is the angular frequency of the acoustic wave.

The acoustic propagation constant can be written in the form

$$\frac{k^2}{\omega^2} = \left(\frac{1}{c} - \frac{i\alpha}{\omega}\right)^2 = \frac{\rho_0 \kappa_T}{\gamma_{\text{eff}}}, \tag{6}$$

where c is the acoustic velocity, α is the attenuation, ρ_0 is the equilibrium density, κ_T is the compressibility, and

$$\gamma_{\text{eff}} = \frac{(C_v)_{\text{eff}} + R}{(C_v)_{\text{eff}}} \tag{7}$$

with R the gas constant. For this simple single relaxation, assuming $\alpha/\omega \ll 1/c$,[6]

$$\alpha\lambda = \pi\left(\frac{c}{c_0}\right)^2 \epsilon \frac{\omega\tau_s}{1 + (\omega\tau_s)^2} \tag{8}$$

and

$$\left(\frac{c_0}{c}\right)^2 = 1 - \frac{\epsilon\omega^2\tau_s^2}{1 + \omega^2\tau_s^2}, \tag{9}$$

where $\epsilon = (c_\infty^2 - c_0^2)/c_\infty^2$, λ is the wavelength, c_0 is the speed of sound for $\omega\tau_s \ll 1$, and c_∞ is the speed of sound at frequencies much greater than the relaxation frequency. The adiabatic relaxation time τ_s is related to the isothermal relaxation time τ used earlier by

$$\tau_s = \frac{C_v + R}{(C_v^\infty + R)\tau}. \tag{10}$$

The relaxation frequency f_r, defined as the frequency at which the maximum absorption per wavelength occurs, is related to τ_s by

$$f_r = \frac{c_\infty}{c_0(1/2\pi\tau_s)}. \tag{11}$$

Figure 1 shows typical curves for absorption per wavelength and velocity dispersion due to a single relaxation process. The example here is Fl_2 at 102°C. The figure compares curves representing the above theory with measured values.[7]

In the case of polyatomic gases or mixtures of relaxing diatomic gases, the different relaxing modes can be coupled together by vibration-to-vibration exchanges. Such complex or multiple relaxation processes exhibit the general behavior given by Eqs. (8) and (9), but the magnitude of the absorption and dispersion and the relaxation frequencies can take on new meaning. In this case of multiple relaxing internal energy modes, Eqs. (8) and (9) take on the form[8]

$$\left(\frac{c}{c^\infty}\right)^2 = 1 + \sum_j \frac{\delta_j k_s/k_s^\infty}{1 + (\omega\tau_{s,j})^2}, \tag{12}$$

$$\alpha\lambda\left(\frac{c}{c^\infty}\right)^2 = -\pi \sum_j \frac{\delta_j k_s/k_s^\infty}{1 + (\omega\tau_{s,j})^2}\, \omega\tau_{s,j}, \tag{13}$$

where $\delta_j k_s/k_s^\infty$ is a relaxing adiabatic compressibility (negative) and j indicates that there might be more than one relaxation process. In these complex cases, $\tau_{s,j}$ can no longer be related to a single energy transfer reaction and $\delta_j k_s/k_s^\infty$ can no longer be related to relaxing energy of a specific mode. Instead, the various modes and reaction pathways are coupled. The sums in Eqs. (12) and (13) are over eigenvalues of the energy transfer matrix which simultaneously accounts for all reactions. Equations (12) and (13) have been used to calculate the sound absorption in moist air as a function of frequency and temperature. The standard for such calculations is now based upon these equations. The process is explained in Chapter 32.

Not only can Eqs. (12) and (13) be used to calculate sound absorption and velocity, but the reverse process is also possible, that is, the transition rates can be extracted from measured values of absorption and velocity. However, the number of possible relaxation paths multiplies rapidly with the increase in the number of relaxing modes, and the identification of specific rates becomes a tedious process and usually involves some assumptions. It has been done for only a few special cases. The next section discusses one example, that is, SO_2/O_2 mixtures.

So far, only relaxation processes that involve the exchange of energy in gases have been considered. Conceptually, relaxation processes in liquids result in similar relaxation equations but there are important differences in detail. These differences are brought about by the greater density of molecules that make multibody interactions typical. In this case, the idea of a rate equa-

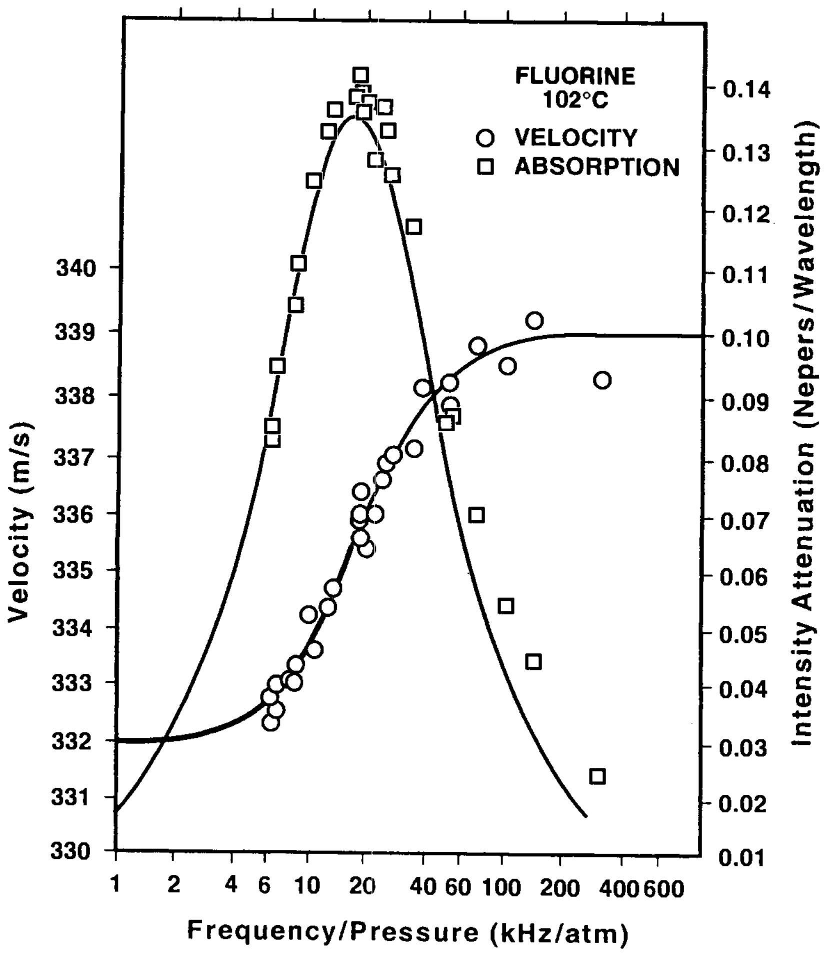


Fig. 1 Sound absorption and dispersion in fluorine at 102°C. The curves are calculated from specific heat data using a value of 17 kHz/atm for the frequency-to-pressure ratio when maximum absorption occurs.

tion is less precise, but the existence of a relaxation time as a measure of the time required for the system to return to equilibrium following a perturbation remains valid.

Various interesting relaxation processes occur in liquids. In some cases, such as CS_2 and a number of organic liquids, the relaxation mechanism seems to be the same as that described above for gases; that is, the internal energy of the individual molecules is being excited in "collisions." Such liquids generally have a positive temperature coefficient of absorption and are called Kneser liquids.[9, 10] It is more common, however, that in liquids molecules temporarily bond to form large groups that may reconfigure themselves when an ultrasonic wave passes through. Such configuration relaxations tend to be very fast and the frequency dependence of the absorption and dispersion may indicate a distribution of relaxation times. Water is an exmaple of such a liquid, generally referred to as associated liquids.[11]

A further complication arises from chemical reactions. For a reversible chemical reaction with heat of reaction ΔH, ΔH enters into the relaxation equations in a manner similar to ΔE for vibrational relaxation. A major difference is that chemical reactions allow the possibility that the number density of molecules can change. Such changes bring about additional relaxation absorption and dispersion.

3 EXPERIMENTAL TECHNIQUES

Before modern techniques were available for measuring transients, most ultrasonic absorption and velocity measurements were made with ultrasonic interferometers.[12–15] This instrument continues to be used with refined precision and modern methods of control and measurement.[16] Both velocity and absorption can be measured with the interferometer. A column of gas or liquid of an adjustable length forms the load on a quartz crystal vibrating at its resonant frequency. The loading effect of the gas or liquid column increases whenever the length of the gas column is a whole multiple of half-wavelengths of sound. This loading effect is reflected in the driving circuit of the crystal. The separation between the peak values determines the sound wavelength and the variation of the magnitude of the peaks with length of the column allows the determination of the absorption coefficient of the sound in the test medium.

Another continuous-wave method for measuring velocity and absorption in gases uses a source and a receiver whose separation can be varied that are mounted in a tube so as to avoid standing waves.[17,18] This traveling-wave tube has been used to make measurements at audible frequencies and reduced pressures and therefore is capable of measurements over the f/p range where many of the interesting relaxation processes occur in gases and gas mixtures. The absorption is determined from the decrease in sound amplitude at the receiver as the source–receiver separation is increased. The wavelength of the sound is equal to the distance between points of equal phase in the received sound. This change in phase is easily observed by comparing the receiver and source signals. With the development of the capability for generating and measuring tone bursts, it has been possible to use the pulse–echo technique.[19,20] In this case the sound velocity is determined from the time of flight and the absorption from the variation of the tone burst amplitude with source and receiver separation.

At lower frequency-to-pressure ratios resonant tubes are used.[3] The velocity is determined from the value of the resonant frequency and the absorption from the width of the resonant peak.

In the case of liquids, special techniques have been developed to measure absorption and velocity at very high frequencies where many of the interesting relaxation effects occur.[21]

When relaxation effects are studied in gases the measurements of sound absorption are designed to provide absorption as a function of frequency divided by pressure over a range of values where the period of the sound wave is approximately equal to the relaxation time. When studying relaxation in molecular N_2, this requires measurements of very small absorption (less than 1 dB/m) at f/p values as low as a fraction of a hertz per atmosphere. On the other hand, measurements of absorption in UF_6 involve very large attenuation (~100 dB/m) at f/p values as high as 10 MHz/atm.[22]

To separate the effects due to vibrational relaxation, the measured absorption must be corrected for the following:

1. Viscous and thermal losses in the body of the gas
2. Viscous and thermal losses to the measuring chamber walls
3. Radiation or leakage losses from the chamber
4. Spreading losses
5. Losses due to rotational relaxation

These different effects are generally small enough to be additive. The quantity usually reported is the absorption coefficient α, in terms of which the amplitude of the plane wave is written as

$$A = A_0 e^{-\alpha(x_2 - x_1)}. \tag{14}$$

Plotting log A versus x gives α as a slope in nepers per metre. The measured α is due to a combination of the losses listed above. Careful design of the experimental apparatus is necessary if accurate corrections are to be made for wall and/or spreading losses.

A common geometry for sound absorption measurements is a cylindrical tube. This method generally avoids losses 3 and 4 (due to spreading and leakage) above and involves well-known corrections for 1 and 2. In this case, the total absorption is given by

$$\alpha_{\text{tot}} = \alpha_{\text{cl}} + \alpha_{\text{rot}} + \alpha_{\text{tu}} + \alpha_v, \tag{15}$$

where α_{cl} (classical absorption) is the absorption due to viscosity and thermal conductivity, α_{rot} is absorption due to relaxation processes not of interest to the study (rotational relaxation for example), α_{tu} is the tube absorption, and α_v is the absorption due to the relaxation process of interest. The term α_{cl} can be computed from known or measured values of viscosity and thermal conductivity[23]; it is proportional to f^2. The term α_{rot} generally is estimated based upon other studies. At frequencies well below the rotational relaxation frequency, it also is proportional to f^2. The tube absorption α_{tu} depends upon the viscosity, thermal conductivity, and tube radius (r). It varies approximately as[24] $[f/(pr^2)]^{1/2}$.

Only plane waves will be present in the tube if the sound frequency is maintained below the cutoff frequency for the first nonplane mode, that is, $f < 0.586c/d$, where c is the sound speed and d the tube diameter. Plane waves can be maintained for higher frequencies if the transducer generating the waves fills the tube and is carefully maintained perpendicular to the tube axis and its surface vibrates as a piston.[25] Interferometers used for high-frequency measurements in liquids and gases will generally satisfy these conditions. As the wavelength becomes small compared to the tube diameter, wall losses become negligible.

Low-frequency absorption measurements in liquids are difficult. Because of the high sound velocity and small absorption, large volumes are required. If the measurements are made in an unbounded liquid, spreading losses will likely dominate. If they are made in an enclosure, the small compressibility of the liquid makes the rigid-wall assumption invalid and it is difficult to correct for wall losses.

In the above discussion of relaxation in gases the relaxation absorption and dispersion of sound resulted when a finite time was required for the passage of the energy of translation into the vibrational energy of the gas molecules. Relaxation mechanisms can also be studied in special cases using the reverse processes. The laser light of a particular frequency is used to excite a vibrational mode. This excess vibrational energy then relaxes into translational energy generating an increase in translational temperature and pressure. If the laser light is chopped, a sound wave is generated. This phenomenon is referred to as the *optoacoustic effect* and the device as a *spectrophone*.[26] The relaxation time can be determined from the phase relations between modulated laser light and the resulting sound pressure.

The primary attraction of optoacoustic measurements is the ability to excite a specific internal mode and observe how energy from that mode makes its way to translation. Experimentally, only a limited number of internal modes can be excited with available lasers, which limits the list of systems that can be studied.

4 TYPICAL RESULTS

A comprehensive review of all studies of ultrasonic relaxation is far beyond the scope of this chapter. No such review has been published in recent years. The best available is that by Herzfeld and Litovitz, which is now 37 years old.[1] A large compilation of experimentally measured absorption and dispersion is provided in Ref. 27. For the purposes of this chapter specific studies will be selected that illustrate the physics involved. Three cases will be treated. The first is the halogen family of diatomic molecules, specifically F_2, Cl_2, Br_2, and I_2. This

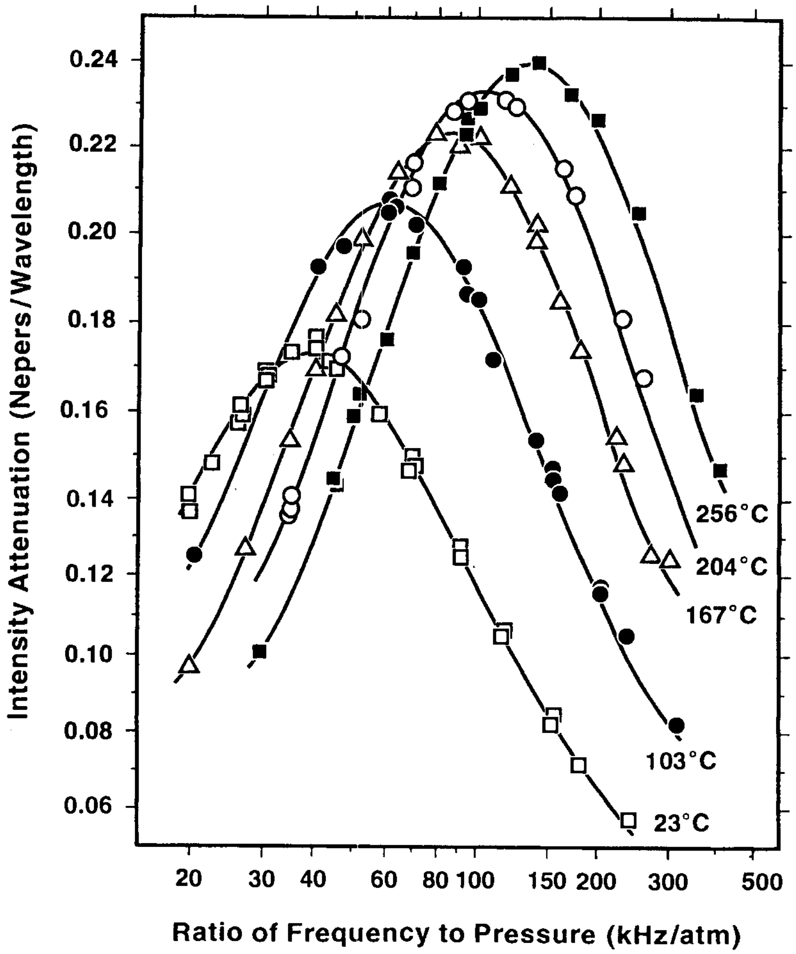


Fig. 2 Relaxation absorption coefficient per wavelength vs. log (frequency/pressure) for chlorine. The solid curves are the theoretical absorption with values of A_m and f_m adjusted to give the best fit of the experimental points.

example illustrates the functional forms presented earlier for absorption due to relaxation processes, gives typical values for probability of energy transfer, $P^{1 \to 0}$, and provide some indication of how $P^{1 \to 0}$ varies with temperature, molecular weight, and vibrational frequency. The second example is a more complex molecule, SO_2, in mixtures with Ar and O_2. This example illustrates the complexity of the relaxation process when different vibrational modes exchange energy. The final example will be for CS_2, which illustrates the case where relaxation has been studied in both the liquid and gas phases.

4.1 Simple Relaxation in Halogen Gases

Shields used an acoustic traveling-wave tube similar to that developed by Angona[17] to measure acoustic absorption in F_2, Cl_2, Br_2, and I_2 as a function of frequency, pressure, and temperature.[7,28] Results for Cl_2 at five different temperatures are plotted in Fig. 2. The absorption and velocity in F_2 at 102°C are shown in Fig. 1. The curves drawn through the experimental points were calculated using the theory discussed in Section 2.

Several features of these curves are interesting. First, when more than one point is plotted for the same value of f/p, it means that absorption was measured at two different pressures but at different frequencies so that the ratio was the same. The agreement between such measured values confirms that the classical and tube corrections were being made correctly and that the relaxation absorption varied as f/p. Second, note that the relaxation time decreases as temperature increases (the relaxation frequency increases). This means that although the density is lower for a given pressure, the probability of energy transfer increases more rapidly than the collision frequency decreases. Finally, note that the maximum absorption increases as the temperature increases. This is a result of increasing vibrational specific heat. The Planck–Einstein relation predicts that

$$C' = R\left(\frac{\theta}{T}\right)^2 \frac{e^{\theta/T}}{(e^{\theta/T} - 1)^2}, \qquad (16)$$

where C' is the vibrational specific heat, θ is the vibrational temperature, and T is the gas temperature.

Results of Shields's study are summarized in Fig. 3, which shows the probability for deexciting the first excited vibrational level in a collision. The log of this probability is plotted versus $T^{-1/3}$. This type of plot was suggested by an early theory by Landau and Teller.[5] They predicted an approximately linear relationship between log P^{10} versus $T^{-1/3}$. From similar measurements on a great many other gases one can expect the following trends for vibration-to-translation transitions:

1. The log of the transition probability increases roughly linearly with $T^{-1/3}$ for a particular molecular collision pair.

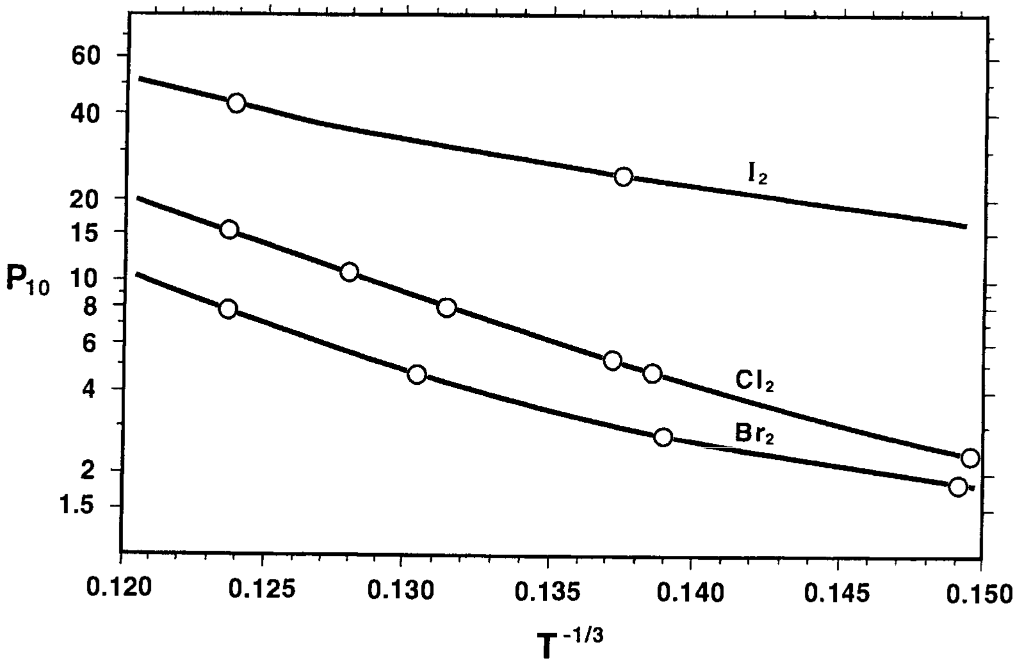


Fig. 3 Log of collision efficiency vs. absolute temperature to the minus-one-third power. The values of the ordinate should be multiplied by 10^{-5} for Cl_2 and by 10^{-4} for Br_2 and I_2.

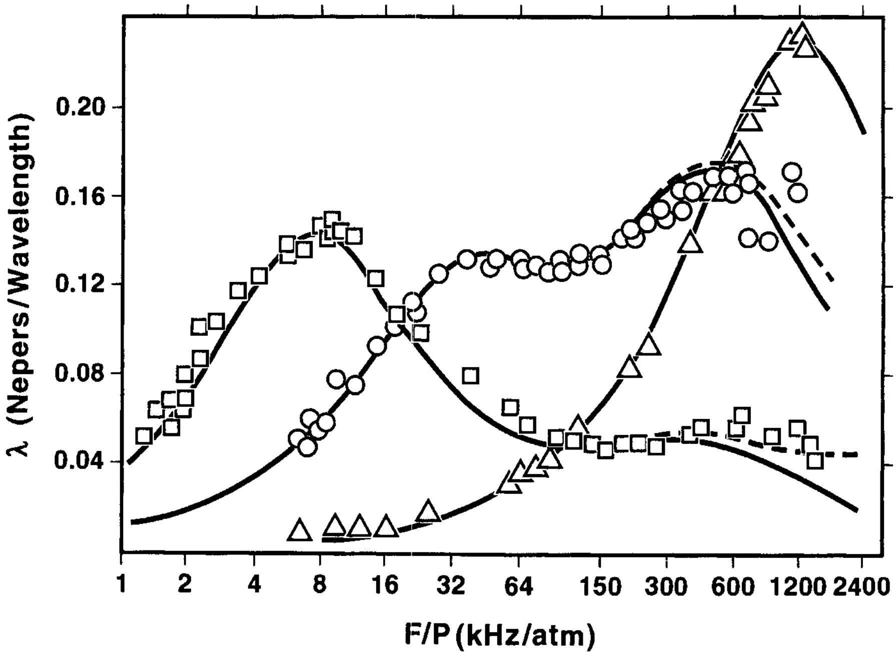


Fig. 4 Relaxation absorption in 75% SO_2/25% O_2 (Δ), 20% SO_2/80% O_2(0), and 5% SO_2/95% O_2 (□) at 500 K. The solid curve was computed assuming series relaxation and the transmission probabilities in Refs. 34–37. From Ref. 34.

2. The transition probability is very sensitive to the amount of energy that must be transferred between vibration and translation in the transition, increasing rapidly as this energy decreases.
3. The transition probability is very sensitive to the time involved in the molecular collision, increasing rapidly as this time decreases. What this means is that the transition probability increases with increasing temperature (see 1 above) and with decreasing molecular mass.
4. For polyatomic molecules, bending vibrations are more easily excited than stretching vibrations (as one might expect from the geometry of the collision process).
5. Water vapor and a few other molecules with low moments of inertia and, therefore, high rotational velocities are very efficient in shortening relaxation times when added to gases, even in small amounts. For such mixtures the relaxation time is much less sensitive to temperature change and can even increase with temperature. This effect is now attributed to the coupling between vibration and rotation in such collisions.[29–33]

4.2 Relaxation in a Polyatomic Molecule

As an example of the use of sound absorption measurements to determine the relaxation scheme in a polyatomic gas mixture, consider the case of SO_2–O_2 mixtures.[34] The SO_2 has three vibrational modes. These can be classified as a bending mode with a fundamental vibrational frequency of 518 cm^{-1}, a symmetrical stretching mode, and an asymmetrical stretching mode with frequencies of 1151 and 1361 cm^{-1}, respectively. Oxygen has a vibrational frequency of 1580 cm^{-1}. Sulfur dioxide was one of the first gases in which the sound absorption versus frequency curve evidenced more than one relaxation time. Figure 4 shows measurements in three different SO_2–O_2 mixtures at 500 K. The solid curves were calculated using the theory of Section 2. Twelve different transition rates were adjusted to make the theoretical curves simultaneously fit these data and the data for pure SO_2, SO_2–Ar mixtures,[35,36] and pure O^2.[37]

As an illustration of the kind of information that can be obtained from such studies, the following conclusions from the SO_2–O_2 and SO_2–Ar measurements are listed:

1. The relaxation in SO_2 is primarily a series process where translational energy flows into the bending mode and from there is shared with the symmetrical stretch with a two quantum for one exchange and from there with the asymmetrical stretch with a one-for-one exchange.
2. Argon is only about one-tenth and O_2 about one-

fourth as effective as another SO_2 molecule in deexciting the bending vibration in SO_2.

3. The first excited level of the O_2 vibration exchanges energy in vibration-to-vibration exchange primarily with the first excited level of the asymmetric stretch mode of SO_2.
4. The transition rate for the SO_2 vibration-to-vibration transfer increases with temperature faster than the vibration-to-translation rates. This is a peculiar property of SO_2 and has been attributed to its dipole moment.

4.3 Relaxation in CS_2 Gas and Liquid

Gaseous CS_2 has an unusually high vibrational specific heat at room temperature. From what has been said above, it is expected, therefore, to have a large relaxation absorption at moderate f/p values. Such was found early to be the case.[38,39] Only a single relaxation absorption peak was found, indicating a series relaxation process with energy passing from translation first into the low-energy bending vibration and from there shared quickly with the stretching modes. The frequency of maximum absorption increased from about 370 to 560 kHz/atm as the temperature increased from 0°C to 160°C. This corresponds to a transition probability that decreases relatively slowly (by about one-third) in this temperature interval.[39]

More interesting is the absorption of CS_2 as a liquid. It is an example of a "Kneser" liquid where the vibrational relaxation can be identified in the liquid state just shifted to a higher frequency due to the higher molecular collision rate in the liquid state.

The concept of a collision loses meaning when dealing with liquids since all molecules are interacting with one or more other molecules simultaneously. If one does assume that the liquid is a very dense gas, a collision frequency in the range of 10^{13} s^{-1} would be appropriate. This immediately suggests that relaxation processes in liquids will be fast. In fact, for most liquids up to a frequency of 10 MHz, no relaxation peak is observed so we must assume that $\omega\tau \ll 1$. Invoking this low-frequency limit, Eq. (8) gives

$$\alpha = bf^2, \qquad (17)$$

where b depends upon the relaxation time and relaxation strength. This is the same frequency dependence exhibited by viscous losses. As a result, relaxation, viscous, and thermal conduction losses are difficult to sort out for most liquids.

Carbon disulfide has a maximum in the quantity α/f^2 in the range of 100 MHz.[40] Takagi has shown that the experimental data are best explained in terms of two relaxation processes. The strongest process is associated with relaxation of the two lowest vibrational modes (ν_2 and ν_1) and has a relaxation time of 2.2 ns. The weaker relaxation is associated with relaxation of ν_3 with a relaxation time of 30 ns.

The observation that relaxation in CS_2 behaves like thermal relaxation in a dense gas makes it a very interesting liquid. Carbon disulfide also strongly absorbs 337 nm UV, which makes it amenable to study with a pulsed N_2 laser.[41] Optoacoustic studies using a pulsed N_2 laser give relaxation times for the slow process of 90 ns. Given the difficulties in extracting the relaxation time for this weak process, a factor of 3 difference should not be surprising.

REFERENCES

1. K. F. Herzfeld and T. A. Litovitz, *Absorption and Dispersion of Ultrasonic Waves*, Academic, New York, 1959.
2. T. G. Winter and G. L. Hill, "High-Temperature Ultrasonic Measurements of Rotational Relaxation in Hydrogen, Deuterium, Nitrogen, and Oxygen," *J. Acoust. Soc. Am.*, Vol. 42, No. 4, 1967, p. 848.
3. A. J. Zuckerwar and W. A. Griffin, "Resonant Tube for Measurement of Sound Absorption in Gases at Low Frequency/Pressure Ratios," *J. Acoust. Soc. Am.*, Vol. 68, No. 1, 1980, p. 218.
4. G. L. Hill and T. G. Winter, "The Effect of Temperature on the Rotational and Vibrational Relaxation Times of Some Hydrocarbons," *J. Chem. Phys.*, Vol. 49, 1968, p. 440.
5. L. Landau and E. Teller, "Zur Theorie der Schalldispersion," *Phys. Z. Sowjet-Union*, Vol. 10, 1936, p. 34.
6. R. T. Beyer and S. V. Letcher, *Physical Ultrasonics*, Academic, New York, 1969, p. 103.
7. F. D. Shields, "Thermal Relaxation in Fluorine," *J. Acoust. Soc. Am.*, Vol. 34, No. 3, 1962, p. 271.
8. H. E. Bass, L. C. Sutherland, J. Piercy, and L. Evans, "Absorption of Sound by the Atmosphere," *Phys. Acoust.*, Vol. XVII, 1984, p. 161.
9. R. T. Beyer and S. V. Letcher, *Physical Ultrasonics*, Academic, New York, 1969, Chap. 5.
10. K. F. Herzfeld and T. A. Litovitz, *Absorption and Dispersion of Ultrasonic Waves*, Academic, New York, 1959, Chap. 8.
11. K. F. Herzfeld and T. A. Litovitz, *Absorption and Dispersion of Ultrasonic Waves*, Academic, New York, 1959, Chap. 12.
12. J. C. Hubbard, "The Acoustic Resonator Interferometer: I. The Acoustic System and Its Equivalent Electric Network," *Phys. Rev.*, Vol. 38, 1931, pp. 1011–1019.
13. J. V. Connter, "Ultrasonic Dispersion in Oxygen," *J. Acoust. Soc. Am.*, Vol. 30, 1958, p. 297.
14. E. S. Stewart and J. L. Stewart, "Rotational Dispersion

in the Velocity, Attenuation, and Reflection of Ultrasonic Waves in Hydrogen and Deuterium," *J. Acoust. Soc. Am.*, Vol. 24, 1952, p. 194.

15. R. T. Lagemann, "The Use of Glass Parts in Ultrasonic Interferometers," *J. Acoust. Soc. Am.*, Vol. 24, 1952, p. 86.
16. B. Jacobs, R. Cerami, J. R. Olson, and R. C. Amme, "Vibrational Relaxation in SiF_4 Using a Computer Controlled Ultrasonic Interferometer," *J. Acoust. Soc. Am.*, Vol. 88, 1990, pp. 2812–2815.
17. F. A. Angona, "Attenuation of Sound in a Tube," *J. Acoust. Soc. Am.*, Vol. 25, 1953, p. 336.
18. F. D. Shields, "Measurements of Thermal Relaxation in CO_2 Extended to 300°C," *J. Acoust. Soc. Am.*, Vol. 31, 1959, p. 248.
19. H. E. Bass, T. G. Winter, and L. B. Evans, "Vibrational and Rotational Relaxation in Sulfur Dioxide," *J. Chem. Phys.*, Vol. 54, 1971, p. 644.
20. F. D. Shields, H. E. Bass, and L. N. Bolen, "Tube Method of Sound-Absorption Measurement Extended to Frequencies Far above Cutoff," *J. Acoust. Soc. Am.*, Vol. 62, 1977, pp. 246–253.
21. K. F. Herzfeld and T. A. Litovitz, *Absorption and Dispersion of Ultrasonic Waves*, Academic, New York, 1959, Chap 9.
22. D. Cravens, F. D. Shields, H. E. Bass, and W. D. Breshears, "Vibrational Relaxation of UF_6: Ultrasonic Measurements in Mixtures with AR and N_2," *J. Chem. Phys.*, Vol. 71, 1979, pp. 2797–2802.
23. F. D. Shields and R. T. Lagemann, "Tube Corrections in the Study of Sound Absorption," *J. Acoust. Soc. Am.*, Vol. 29, No. 4, 1957, p. 470.
24. F. D. Shield and J. Faughn, "Sound Velocity and Absorption in Low-Pressure Gases Confined to Tubes of Circular Cross Section," *J. Acoust. Soc. Am.*, Vol. 46, No. 1, 1969, p. 158; F. D. Shields, "An Acoustical Method for Determining the Thermal and Momentum Accommodation Coefficients of Gases On Solids," *J. Chem. Phys.*, Vol. 62, No. 4, 1975, p. 1248.
25. F. D. Shields, H. E. Bass, and L. N. Bolen, "Tube Method of Sound-Absorption Measurement Extended to Frequencies Far above Cutoff," *J. Acoust. Soc. Am.*, Vol. 62, No. 2, 1977, p. 346.
26. H. E. Bass and H. X. Yan, "Pulsed Spectrophone Measurements of Vibrational Energy Transfer in CO_2," *J. Acoust. Soc. Am.*, Vol. 74, No. 6, 1983, p. 1817.
27. W. Schaafs, "Landolt–Bornstein Numerical Data and Functional Relationships in Science and Technology; Group II; Atomic and Molecular Physics," Vol. 5, *Molecular Acoustics*, Springer Verlag, Berlin, 1967.
28. F. D. Shields, "Sound Absorption in the Halogen Gases," *J. Acoust. Soc. Am.*, Vol. 32, No. 2, 1960, p. 271.
29. J. W. L. Lewis and K. P. Lee, "Vibrational Relaxation in Carbon Dioxide/Water Vapor Mixtures," *J. Acoust. Soc. Am.*, Vol. 38, No. 5, 1965, p. 813.
30. F. D. Shields and J. A. Burks, "Vibrational Relaxation in CO_2/D_2O Mixtures," *J. Acoust. Soc. Am.*, Vol. 43, No. 3, 1968, p. 510.
31. F. D. Shields, "Sound Absorption and Velocity in H_2S and CO_2/H_2S Mixtures," *J. Acoust. Soc. Am.*, Vol. 45, No. 2, 1969, p. 481.
32. F. D. Shields, and G. P. Carney, "Sound Absorption in D_2S and CO_2/D_2S Mixtures," *J. Acoust. Soc. Am.*, Vol. 47, No. 5, 1970, p. 1269.
33. H. E. Bass and F. D. Shields, "Vibrational Relaxation and Sound Absorption in O_2/H_2O Mixtures," *J. Acoust. Soc. Am.*, Vol. 56, No. 3, 1974, p. 856.
34. B. Anderson, F. D. Shields, and H. E. Bass, "Vibrational Relaxation in SO_2/O_2 Mixtures," *J. Chem. Phys.*, Vol. 56, No. 3, 1972, p. 1147.
35. F. D. Shields, "Vibrational Relaxation in SO_2 and SO_2/Ar Mixtures," *J. Chem. Phys.*, Vol. 46, No. 3, 1967, p. 1063.
36. F. D. Shields and B. Anderson, "More on Vibrational Relaxation in SO_2/Ar Mixtures," *J. Chem. Phys.*, Vol. 55, No. 6, 1971, p. 2636.
37. F. D. Shields and K. P. Lee, "Sound Absorption and Velocity Measurements in Oxygen," *J. Acoust. Soc. Am.*, Vol. 35, No. 2, 1963, p. 251.
38. W. T. Richards and J. A. Reid, "Acoustical Studies, IV. The Collison Efficiences of Various Molecules in Exciting Lower Vibrational States of Ethylene, Together with Observations Concerning the Excitation of the Rotational Energy in Hydrogen," *J. Chem. Phys.*, Vol. 2, 1934, p. 206.
39. J. C. Gravitt, "Sound Absorption in Carbon Disulfide Vapor as a Function of Temperature," *J. Acoust. Soc. Am.*, Vol. 32, No. 5, p. 560.
40. K. Takagi, "Vibrational Relaxation in Liquid Carbon Disulfide," *J. Acoust. Soc. Am.*, Vol. 71, 1982, pp. 74–77.
41. C. H. Thompson, S. A. Cheyne, H. E. Bass, and R. Raspet, "Optoacoustic Observation of Internal Relaxation in Liquid CS_2," *J. Acoust. Soc. Am.*, Vol. 85, No. 6, 1989, pp. 2405–2409.

57

PHONONS IN CRYSTALS, QUASICRYSTALS, AND ANDERSON LOCALIZATION

JULIAN D. MAYNARD

1 INTRODUCTION

The term *phonon* refers to the effects of a transition between eigenstates of a system of coupled quantum mechanical oscillators. The phonon is sometimes referred to as a "quantized sound wave," but phonons and sound waves differ in some ways: phonons usually apply to discrete, strictly linear systems, while sound waves are usually derived from continuous, intrinsically nonlinear systems in the limit of small amplitudes. The difference appears in the relation between energy and momentum (second-order quantities), but the difference is small in real systems. Phonons may occur in all phases of matter, but they are most readily apparent in crystalline solids. In order to understand the effects of disordered and quasiperiodic systems, it is necessary to know some details of the periodic system. In Section 3 we derive the formulas for a system of harmonically coupled classical particles in one dimension. Section 4 covers a one-dimensional system of harmonically coupled quantum particles. Section 5 discusses the effects of unit cells with more than one particle and the extension of the theory to two and three dimensions. Section 6 treats the effects of disorder in the system and explains Anderson localization. Section 7 discusses the effects of quasicrystalline symmetry.

2 BRIEF HISTORY

The first evidence for phonons in solids occurred in the observation by Dulong and Petit in 1819 that the specific heat of a solid is usually about twice that of the corresponding gas. This suggested that in addition to the kinetic energy that gas molecules possess, solids also have a way of storing potential energy. Calculations of wave motion in periodic lattices were done by Lord Kelvin in 1881, but the relevance to the properties of solids was not known then. An atomistic theory of the specific heat of solids was first proposed by Einstein in 1907. He assumed that the kinetic and potential energies arose from atoms oscillating about their equilibrium lattice positions, and using the newly formulated Planck quantum theory, he related the energy to the frequency of vibration. Einstein made the simplifying assumption that the atoms oscillated independently of each other, and his formula for the specific heat was consequently incorrect. The coupling of the atomic oscillators was properly taken into account by Peter Debye in 1912, and later in more detail by Max Born, Theodore van Karman, and Moses Blackman. Their theory correctly matched the experimental measurements of the temperature dependence of the specific heat of solids. In 1912 Debye recognized the role of phonons in explaining thermal conductivity, and in the same year Frederick Lindemann showed that the lattice vibrations were related to thermal expansion and the melting of solids. These last two effects are due to nonlinearities in the forces between the atoms in the solid.

3 HARMONICALLY COUPLED CLASSICAL PARTICLES

We consider a one-dimensional sequence of masses m_j $(j = 0, N + 1)$ connected with ideal springs of stiffness s_j $(j = 0, N)$. The spring s_j lies between mass m_j and

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker
Copyright © 1997 John Wiley & Sons, Inc.

mass m_{j+1}. The displacement of the mass m_j is denoted u_j; u_0 and u_{N+1} serve as boundary conditions at each end of the system. Inside the boundaries the motion of each mass is governed by Newton's law and Hooke's law:

$$m_j \frac{d^2 u_j}{dt^2} = -s_{j-1}(u_j - u_{j-1}) + s_j(u_{j+1} - u_j). \qquad (1)$$

We assume that the motion can be Fourier analyzed in time, and adopt a time dependence $\exp(i\omega t)$. The first term in Eq. (1) becomes $-m_j\omega^2 u_j$.

At this point standard derivations make the sequence periodic and invoke Bloch's theorem, which provides the form for the normal modes of the coupled system. However, Bloch's theorem is usually derived for an infinitely periodic system or a system that has periodic boundary conditions. In the derivation presented here, a technique will be used that is valid for arbitrary boundary conditions and that will furthermore provide useful insight into the properties of a disordered (nonperiodic) system.

The system of equations represented by Eq. (1) may be written as 2 × 2 matrix equations:

$$\begin{pmatrix} u_{j+1} \\ u_j \end{pmatrix} = \begin{pmatrix} 2\alpha_j & -\beta_j \\ 1 & 0 \end{pmatrix} \begin{pmatrix} u_j \\ u_{j-1} \end{pmatrix}, \qquad (2)$$

where $\omega_j = 2\sqrt{(s_j/m_j)}$, $\alpha_j = (1 + \beta_j)/2 - 2(\omega/\omega_j)^2$ and $\beta_j = s_{j-1}/s_j$.

To proceed, we look for the singular value decomposition of the matrix in Eq. (2); that is, we need to find matrices D_j and inverses D_j^{-1} such that

$$\begin{pmatrix} 2\alpha_j & -\beta_j \\ 1 & 0 \end{pmatrix} = D_j^{-1} \begin{pmatrix} \lambda_j^+ & 0 \\ 0 & \lambda_j^- \end{pmatrix} D_j. \qquad (3)$$

The singular values $\lambda_j^\pm$ are found by solving the secular equation given by setting a determinant equal to zero, as in a standard matrix eigenvalue problem.

An equation, which will later be instructive, for periodic and disordered systems is one that provides the normal modes of the system, expressed as a product of the matrices in Eq. (3):

$$\begin{pmatrix} u_{j+1} \\ u_j \end{pmatrix} = D_j^{-1} \begin{pmatrix} \lambda_j^+ & 0 \\ 0 & \lambda_j^- \end{pmatrix} D_j D_{j-1}^{-1} \begin{pmatrix} \lambda_{j-1}^+ & 0 \\ 0 & \lambda_{j-1}^- \end{pmatrix} D_{j-1},$$
$$\ldots, D_1^{-1} \begin{pmatrix} \lambda_1^+ & 0 \\ 0 & \lambda_1^- \end{pmatrix} D_1 \begin{pmatrix} u_1 \\ u_0 \end{pmatrix}. \qquad (4)$$

We now specialize to the case of a periodic system. We let $m_j = m$, $s_j = s$, $\omega_j = \omega_0 = 2\sqrt{(s/m)}$, $\beta_j = 1$, $D_j = D$, and $\alpha_j = \alpha = 1 - 2(\omega/\omega_0)^2$. Because of the cancellation of DD^{-1} in the product in Eq. (4), we have a simple expression for the displacements u_j:

$$\begin{pmatrix} u_{j+1} \\ u_j \end{pmatrix} = D^{-1} \begin{pmatrix} (\lambda^+)^j & 0 \\ 0 & (\lambda^-)^j \end{pmatrix} D \begin{pmatrix} u_1 \\ u_0 \end{pmatrix}. \qquad (5)$$

The expression for $\lambda^\pm$ for the periodic system is

$$\lambda^\pm = \begin{pmatrix} \alpha \mp i(1 - \alpha^2)^{1/2} & \text{if } 0 < \omega < \omega_0 \\ \alpha \mp (\alpha^2 - 1)^{1/2} & \text{if } \omega > \omega_0 \end{pmatrix}. \qquad (6)$$

Note that if $\omega > \omega_0$, then $|\lambda^+| < 1$ and $|\lambda^-| > 1$. Because of the powers of j in Eq. (5), the larger singular value grows exponentially while the smaller singular value decays; these solutions correspond to evanescent waves, which are usually not regarded as solutions of interest, but they are nevertheless important; they account for the band gaps in periodic systems and for the exponentially localized states that appear in the gaps in disordered systems. If $0 < \omega < \omega_0$, then $\lambda^+ = (\lambda^-)^*$, $|\lambda^\pm| = 1$, and we can write $\lambda^\pm = e^{\pm ika}$. The choice of notation for the phase, ka, will be evident later. From the ω dependence in λ (through α), we can solve for ω in terms of k and obtain the dispersion relation

$$\omega_k = \omega_0 \sin(ka/2). \qquad (7)$$

The subscript k has been added to emphasize the dispersion. Note that $0 < k < \pi/a$. This dispersion relation is illustrated in Fig. 1.

We now consider some limits relevant to a continuous system. We assume that the springs connecting the masses have been stretched to a periodic spacing a with the application of a static tension T, so that $T = sa$. We

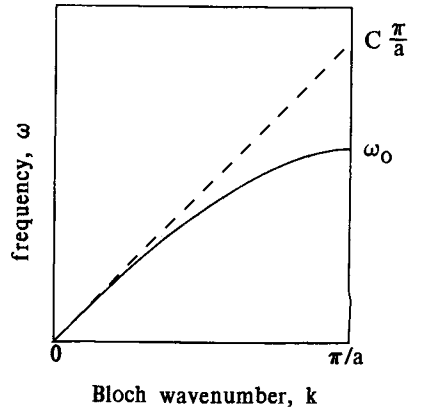


Fig. 1 Dispersion relation for a one-dimensional system of harmonically coupled particles.

now take the limit $m \rightarrow 0$ and $a \rightarrow 0$, while keeping the mass per unit length, $\mu = m/a$, constant. As $a \rightarrow 0$, then $\sin(ka/2) \rightarrow ka/2$, and the dispersion relation becomes $\omega_k = ck$, where $c = \sqrt{(T/\mu)}$ is the usual speed of sound in a string of mass per unit length μ and tension T. Thus the dispersion relation for the continuum limit is obtained.

If one considers the dispersion relation for the discrete system in the limit of small k, one also obtains $\omega = ck$. However, if this is extrapolated to the maximum $k = \pi/a$, one finds a frequency $\omega = c\pi/a$, whereas the actual frequency should be ω_0, which is a factor of $2/\pi$ smaller than the extrapolated value of $c\pi/a$. This dispersion is a fundamental difference between the discrete coupled oscillators appropriate for phonons and sound waves in a continuous system.

In addition to providing the singular values $\lambda^{\pm}$, Eq. (3) is solved for the D matrices. If these and $\lambda^{\pm}$ are inserted into Eq. (5), then the normal modes are obtained:

$$u_j = e^{ik(ja)}U_k + e^{-ik(ja)}U_{-k}, \tag{8}$$

where

$$U_k = \frac{u_1 - u_0 e^{-ika}}{(2i)\sin(ka)}. \tag{9}$$

The values of u_0, u_1, U_k, and restrictions on k (eigenvalues) are determined by the boundary conditions and initial conditions. For example, for a system clamped at the ends, $u_0 = 0$, $u_{N+1} = 0$:

$$u_j = \sin[k(ja)]\left[\frac{u_1}{\sin(ka)}\right], \tag{10}$$

and k is quantized with $k = [n/(N+1)]\pi/a$, $n = 1, 2, \ldots, N$. For an infinite system or a system with periodic boundary conditions one finds $U_k = 0$ or $U_{-k} = 0$. If one uses a coordinate system with the origin at the mass m_0, then the position of the jth mass is $x = ja$, and we can write for infinite or periodic boundary conditions

$$u_k(x) = e^{ikx}U_k, \tag{11}$$

which is the usual Bloch wave result. The subscript k has been added to the displacement field $u(x)$ as a label for the normal mode. The final solution will be a linear combination of the normal modes:

$$u_j = \sum_k (e^{ik(ja)}U_k + e^{-ik(ja)}U_{-k})e^{i\omega_k t}. \tag{12}$$

For any boundary conditions, clamped, periodic, and so forth, one can show that the normal modes are orthogonal and obtain the normalization constants. With this one can derive the expression for the total energy of the system:

$$E = m \sum_k |U_k|^2 \omega_k^2. \tag{13}$$

4 HARMONICALLY COUPLED QUANTUM PARTICLES

For a system of harmonically coupled quantum mechanical particles, we use a Lagrangian formulation instead of the Newton's law formulation in Eq. (1). The Lagrangian is

$$L = \frac{1}{2}\sum_j m_j \left(\frac{du_j}{dt}\right)^2 - \frac{1}{2}\sum_j s_j(u_{j+1} - u_j)^2. \tag{14}$$

The canonical momenta are $p_j = m_j(du_j/dt)$, and the Hamiltonian is

$$H = \sum_j p_j\left(\frac{du_j}{dt}\right) - L. \tag{15}$$

The system is quantized with the commutation relations $[u_j, p_{j'}] = i\hbar\delta_{j,j'}$, where $\hbar$ is Planck's constant divided by 2π. We now assume a periodic system with $m_j = m$ and $s_j = s$. Guided by our results with the classical system, we let

$$u_j = \sum_k (e^{ikja}U_k + e^{-ikja}U_{-k}). \tag{16}$$

The Hamiltonian becomes

$$H = \frac{1}{2m}\sum_k P_k P_{-k} + \frac{1}{2} m \sum_k \omega_k^2 U_k U_{-k}, \tag{17}$$

where $P_k = m(dU_{-k}/dt)$ and ω_k is defined as before in Eq. (7). One could use this Hamiltonian and construct a Schrödinger wave equation for a field $\psi(U)$, where U is a point in the $2N$-dimensional U_k space, and $P_k\psi(U) = -i\hbar(\partial\psi/\partial U_k)$. However, a better approach is to use another transformation and construct the properties of the eigenfunctions and eigenvalues using the commutation relations for u_j and p_j. We define

$$a_k = \left(\frac{m}{2\hbar\omega_k}\right)^{1/2}\left(\omega_k U_k + \frac{i}{m} P_{-k}\right), \tag{18}$$

and $a_k^+ = a^*_{-k}$, with $\omega_{-k} = \omega_k$. The Hamiltonian becomes

$$H = \sum_k \hbar\omega_k\left(N_k + \frac{1}{2}\right), \tag{19}$$

where $N_k = a_k^+ a_k$. Using the commutation relations for u_j and p_j, one finds the commutation relation for a_k^+ and a_k as $[a_k^+, a_{k'}] = \delta_{k,k'}$.

We now write ψ_{kn} for an eigenfunction of the operator N_k with eigenvalue n_k. One can show that the eigenvalues must be positive. Using the commutation relations, one finds that $a_k\psi_{kn}$ is an eigenfunction with eigenvalue $(n_k - 1)$, and $a_k^+\psi_{kn}$ is an eigenfunction with eignevalue $(n_k + 1)$. We then consider the sequence of eigenfunctions given by $(a_k)^p\psi_{kn}$, with p a positive integer. The eigenvalues form the sequences $(n_k - 1), (n_k - 2), \ldots, (n_k - p)$, etc. If this sequence were allowed to continue indefinitely, then for $p > n_k$, the eigenvalues would become negative, which cannot happen. If n_k is such that $(n_k - p) = 0$, then the sequence will terminate without going negative. Thus we have $n_k = p$; that is, n_k is quantized to be a positive integer.

A general state of the system will be a superposition of eigenstates:

$$\psi = \sum_k \sum_n C_{kn}\psi_{kn}, \tag{20}$$

where $|C_{kn}|^2$ gives the probability that the system is in the state ψ_{kn}. The expectation value of the total energy is (using $E\psi = H\psi$)

$$\langle\psi|E|\psi\rangle = \sum_k \sum_n |C_{kn}|^2\hbar\omega_k\left(n_k + \frac{1}{2}\right). \tag{21}$$

By examining the expectation value of the square of the momentum operator, one finds

$$\frac{1}{m}\langle\psi|P^2|\psi\rangle = m\sum_k \langle\psi|U_k|\psi\rangle\omega_k^2 \tag{22}$$

$$= \sum_n \sum_k |C_{nk}|^2\hbar\omega_k n_k = \langle\psi|E|\psi\rangle - E_0, \tag{23}$$

where $E_0 = \sum_k |C_{nk}|^2\hbar\omega_k/2$ is the zero-point energy. If $|C_{nk}|$ is appreciable only for $n_k \gg \frac{1}{2}$, then $\langle\psi|E|\psi\rangle - E_0 \sim \langle\psi|E|\psi\rangle$, and we have from Eqs. (22) and (23)

$$\langle\psi|E|\psi\rangle = m\sum_k \langle\psi|U_k^2|\psi\rangle\omega_k^2, \tag{24}$$

which corresponds to the classical expression in Eq. (13).

If the expectation value of the position operator is examined, it is found that it couples a state ψ_{nk} only to the state $\psi_{(n+1)k}$ or $\psi_{(n-1)k}$. As a consequence, the time dependence will behave like

$$e^{i[(n+1)-n]\omega_k t} + e^{i[(n-1)-n]\omega_k t} = 2\cos(\omega_k t). \tag{25}$$

If $C_{nk} \sim C_{(n\pm1)k}$ and is appreciable for only one k and for $n_k \gg \frac{1}{2}$, then

$$\langle\psi|U_k|\psi\rangle = \left(\frac{2\langle\psi|E|\psi\rangle}{m\omega_k^2}\right)^{1/2}\cos(\omega_k t), \tag{26}$$

which corresponds to the normal mode oscillating between the classical turning points at the frequency ω_k. It is interesting to note that in order to observe a "classical sound wave," the quantum system must not only have a large n_k so that $n_k\hbar\omega_k$ corresponds to the macroscopic classical energy, the quantum state must contain a superposition of two eigenstates, ψ_{nk} and $\psi_{(n\pm1)k}$.

The term *phonon* can now be defined in the following sense: We can say that a phonon is emitted or absorbed when a system of harmonically coupled quantum particles makes a transition from a state ψ_{nk} to a state $\psi_{(n-1)k}$ or $\psi_{(n+1)k}$.

5 LARGER UNIT CELLS AND HIGHER DIMENSIONS

Instead of having a periodic system such that $m_j = m$, and so on, one may have different masses or spring constants but still have a periodic system. That is, several different masses and springs may be joined to form a unit cell, and this unit cell may be repeated periodically. For M different masses or springs, the matrix in Eq. (2) will enlarge from 2×2 to $2M \times 2M$, and instead of a single dispersion relation as in Eq. (7), there will be M dispersion relations and M "branches" in the eigenvalue spectrum. If the system is extended into two or three dimensions, the matrix representing the dynamics of the system is further enlarged, the displacement fields become vector fields, and the eigenfunctions may be characterized as transverse or longitudinal. A thorough discussion of the dynamics of a three-dimensional system of coupled oscillators is available in Ref. 1.

5.1 Phonon Density of States

In statistical physics, one calculates the properties of a system by taking a product of the quantity of interest with a distribution function and integrating over all possible states. The distribution functions are functions of the energy of the state, and for this reason it is usually easier to transform the integration variable from the state label, such as the Bloch wave vector k in our one-dimensional example, to the energy, which is related to ω in our example. In our one-dimensional example, the transformation is accomplished using

$$dk = \left(\frac{d\omega}{dk}\right)^{-1} d\omega = \frac{d\omega}{v_g}, \tag{27}$$

where $v_g = d\omega/dk$ is the group velocity. For the general case in higher dimensions, the Jacobian of the transformation is the density of states, which also varies as $1/v_g$. Values of the state parameter (e.g., k) where the dispersion relation ω_k is flat are called critical points, and at these points the group velocity is zero and the density of states is singular. Such singularities are called van Hove singularities.[1,2] In our one-dimensional example,

$$\frac{d\omega}{dk} = \frac{2\omega_0}{a} \cos\frac{ka}{2} = \frac{2}{a}(\omega_0^2 - \omega^2)^{1/2} \tag{28}$$

so that $d\omega/dk = 0$, and the density of states is singular, when $\omega = \omega_0$. In two and three dimensions the singularities may be classified according to the nature of $|\nabla_k \omega| = 0$, whether it is a maximum, minimum, or one of two types of saddlepoint. The van Hove singularities are sometimes important in determining the properties of a material relating to phonons.

6 EFFECTS OF DISORDER: ANDERSON LOCALIZATION

6.1 Introduction

If one takes the modulus of an eigenfunction for a periodic system [e.g., Eq. (11)], then one obtains a constant, independent of position in the system. That is, the eigenfunction has the same amplitude throughout the entire system. Such eigenfunctions are said to be extended. Extended eigenfunctions are very important in solid-state theory; for example, they explain why metallic crystals have high electrical conductivity. In a metal, an electron is very strongly scattered by the positive ions, and on this basis a metal should have a very low electrical conductivity. But because the ions in a crystal are arranged periodically, the quantum mechanical eigenfunctions for the electron are extended, allowing the electron to propagate freely and resulting in high electrical conductivity.

That the eigenfunctions for a periodic system are extended is guaranteed by Bloch's theorem. We now consider what happens if we take a periodic system and introduce some disorder throughout the system. One would suppose that the eigenfucntions, instead of having uniform amplitude everywhere in the system, would simply obtain some variations about the original constant value, but would otherwise remain extended. Surprisingly, this is not what happens; instead, the eigenfunctions become exponentially localized, having a maximum value at some site in the system and, at least in a probabilistic sense, decaying exponentially from that site. This phenomenon, referred to as Anderson localization, was used by P. W. Anderson and Sir Neville Mott to explain the electrical conductivity of disordered metals, and this accomplishment was cited in their Nobel prize in 1977.

While much of the research with Anderson localization has dealt with electron Schrödinger waves in disordered metals, the phenomenon is relevant for any form of wave propagation in a disordered system. Anderson localization is important for the understanding of phonon quantum wave propagation and its effects on specific heat and thermal conductivity,[3] and for the understanding of classical wave propagation in disordered systems, as in acoustic geophysical survey, oil exploration, medical imaging, the vibration of complex structures, sound propagation in the ocean, radar propagation in vegetation, and so forth. Anderson localization remains an active area of research for both quantum and classical systems in condensed matter physics.

Many of the advances in the theory of Anderson localization have occurred through the application of renormalization group methods and other modern techniques of statistical physics. As with critical phenomena in statistical physics, Anderson localization is strongly dependent on the dimensionality of the system. In one dimension it is found that any amount of disorder will result in some degree of localization; very weak disorder will result in localization, but the characteristic length for the exponential decay may be large. In three dimensions it is believed that a critical amount of disorder must be exceeded before all of the eigenstates become localized. Two dimensions is the "critical dimension," where theory becomes difficult; it is believed that in two dimensions all eigenstates are localized, as in one dimension, but the localization is much weaker. In one dimension there are rigorous theorems that describe Anderson localization, but in two and three dimensions there are no comprehensive rigorous theorems. In the paragraphs below Anderson localization in one dimension, based on the coupled oscillator model described in the first sections of this chapter, will be treated. More detailed reviews of Anderson localization, particularly in dimensions higher than one, may be found in Refs. 4–6.

6.2 Explanation of Anderson Localization

While it is easy to describe Anderson localization, it is very difficult to understand how it occurs; one would naively expect that disorder, especially weak disorder, would simply put variations in the amplitude of the otherwise extended eigenfunctions. Of the many studies treating Anderson localization, most simply assume the phenomenon as an established fact and few offer explanations. The rigorous theory for one dimension was found by Furstenburg in 1963[7]; Furstenburg's study is long and difficult, with references to earlier theorems, and so it offers little physical insight into the phenomenon. In this section, a simple but nonrigorous explanation of Anderson localization will be presented. An excellent, but longer, explanation may be found in a study by Luban and Luscombe.[6]

A careful statement of the Anderson localization effect reveals some of the difficulty. Anderson localization does not refer to the behavior of a single realization of a disordered system, but rather refers to the results of averaging over an ensemble of realizations, with the disorder of each realization randomly determined with some distribution function. Furthermore, the rigorous result occurs in the limit as the system becomes infinite in size. The actual eigenfunctions of a single realization do not have the form $\exp(-2|x-x_l|/\Lambda)$ (here x_l would be the site of localization and Λ would be the localization length); in an actual realization, an eigenfunction at some site far from x_l may fluctuate above the small exponential tail, requiring an unbounded prefactor to the exponential. When averages are taken, the fluctuations can be ignored, and the eigenfunctions, while not actually having the form $\exp(-2|x - x_l|\Lambda)$, behave as though they did, in some statistical sense.

In the treatment of the coupled oscillators presented in the first sections, it was found that the eigenfunctions could be expressed as products of matrices, as in Eq. (4). For an ensemble of disordered systems, we consider a product of random matrices for a single realization, which we write symbolically as

$$P_j = \prod_{j'=1}^{j} M_{j'}. \tag{29}$$

We are interested in the ensemble average $\overline{P}_j$ of the product of random matrices. While this is a difficult mathematical problem, we consider an analogy with a much more familiar problem, the ensemble average of a sum of random variables $X_{j'}$ with distribution function $F(X)$:

$$S_j = \sum_{j'=1}^{j} X_{j'}. \tag{30}$$

We expect from intuition, and know from rigorous theorems, that the average of the sum is just the average of the random variable, $\overline{X} = \int XF(X)dX$, times j, so that $\overline{S}_j = j\overline{X}$ in the limit as j goes to infinity. Comparing Eq. (29) with Eq. (30) we imagine taking the log of the product of random matrices and obtain a sum of random variables:

$$\ln(P_j) = \sum_{j'=1}^{j} \ln(M_{j'}). \tag{31}$$

We might then expect the average of $\ln(P_j)$ to be j times some mean $\overline{M}$, so that $\overline{\ln(P_j)} = j\overline{M}$. Exponentiating both sides of this expression suggests that $\overline{P}_j = \exp(j\overline{M})$. This indicates that the ensemble average of a product of random matrices may grow or decay exponentially as j goes to plus or minus infinity. This effect might also be seen in Eq. (4), showing a product of random matrices, including their singular value decomposition. If the product in Eq. (4) is ensemble averaged, one might expect that the combinations DD^{-1} would tend to cancel; the diagonal matrices cannot cancel, leaving a matrix of the form:

$$P_j = \begin{pmatrix} \prod_{j'=1}^{j} \lambda_{j'}^{+} & 0 \\ 0 & \prod_{j'=1}^{j} \lambda_{j'}^{-} \end{pmatrix} \tag{32}$$

For the periodic system we would have $|\lambda_{j'}^{\pm}| = 1$, resulting in extended eigenfunctions. However, for the disordered system we have $|\lambda_{j'}^{-}| > 1$ and $|\lambda_{j'}^{+}| < 1$, so that in Eq. (32) the singular values (and the eigenfunctions) grow or decay exponentially.

These discussions of the average of a product of random matrices are not rigorous, but they do suggest what Furstenburg proved rigorously. The statistical nature of the Anderson localization is actually stated somewhat differently using Furstenburg's theorem: For waves propagating in a disordered system, the eigenfunctions decay exponentially to zero at plus or minus infinity with probability one. This probabilitistic statement permits actual wave functions for a given realization to fluctuate above an exponential envelope.

The problem with the rigorous statement of Anderson localization is that it is difficult to see how to apply it to practical situations. That is, real experiments may have only one finite-size realization available; one cannot examine the system out to infinity: even if one could, it would not be evident if something were happening with "probability one." A treatment of Anderson localization that illustrates the localization and its probabilistic nature in a single finite-size realization is the one by Luban and Luscombe.[6] In this treatment one considers the effects of the rigorous theorems on a computer evaluation of a finite product of random matrices. For a complete understanding of Anderson localization, the reference by Luban and Luscombe[6] is highly recommended.

The most common description of Anderson localization, namely that "the eigenfunctions are exponentially localized" is not particularly useful for single realizations of finite-size systems. The eigenfunctions only behave as though they were exponentially localized if many systems are ensemble averaged. A better way of describing the effects of disorder is with a quantity called the participation ratio, $P(N)$ [or $P(L)$], which is defined with

$$P(N) = \frac{\left(\sum_0^N |y_j|^2\right)^2}{\left(\sum_0^N |y_j|^4\right)} \tag{33}$$

for eigenfunctions of discrete systems or

$$P(L) = \frac{\left(\int_{-L/2}^{L/2} |\psi(x)|^2 dx\right)^2}{\left(\int_{-L/2}^{L/2} |\psi(x)|^4 dx\right)} \tag{34}$$

for continuous systems. The participation ratio is a measure of the localization of a function: A function that is spread out has a larger value of $P(L)$ than a function that is more localized. If $|\psi(x)| = \psi_0$ then $P(L) = L$, that is, the entire system participates. If $|\psi(x)| = \psi_0 \exp(-2|x|/\Lambda)$, then $P(L) \rightarrow \Lambda$, the localization length, as $L \rightarrow \infty$. A practical method of determining whether a particular eigenfunction is localized or not is to monitor $P(L)$ as L is increased beyond an anticipated localization length Λ. If $P(L)$ keeps increasing as L is increased, then the state is not localized; if $P(L)$ levels off at a finite value as L is increased, then the state is localized. In either case $P(L)$ may fluctuate about L or Λ.

7 QUASICRYSTALLINE SYMMETRY

Until a few years ago it was believed that solid matter could exist in two basic forms: crystalline and amorphous (glassy). In an amorphous material the atoms are positioned at random positions, and a macroscopic sample would be homogeneous and isotropic. Crystals are formed by taking a unit cell and periodically repeating it to fill three-dimensional space. In order to fill all space with long-range periodic order, only unit cells of particular shapes are allowed; only 14 shapes fit together to fill space, and these form the bases for the 14 Bravis lattices of crystallography. Since there are only 14 possible unit cells, only certain rotational symmetries are allowed; in particular, fivefold rotational symmetry is forbidden. However, a few years ago Shechtman[8] made an aluminum alloy whose x-ray diffraction pattern indicated fivefold rotational symmetry. While it is possible to have atoms in clusters with fivefold (icosahedral) short-range order, the diffraction patterns had sharp spots, indicating long-range order. At first the existence of such materials seemed impossible, but theorists pointed out that while it is impossible to have fivefold rotational symmetry and long-range periodic order, it is possible to have fivefold rotational symmetry and another type of long-range order that is quasiperiodic. Thus atoms may minimize their energy by arranging themselves quasiperioidically, forming quasicrystals.

In one dimension the notion of quasiperiodicity is relatively easy to understand. For a line with periodic lattice sites at a spacing a, the Fourier transform (or diffraction pattern) has a sharp line at π/a. If a line has points positioned randomly, then the Fourier transform would show a broad spectrum. Now suppose one takes a line with periodic lattice spacing b, and superimposes it on the line with periodic spacing a. The result may appear similar to the line with random spacing. However, if the two lattice constants a and b are commensurate, that is, their ratio a/b is equal to the ratio of two integers, then the superposition will be periodic. The period may be much larger than a or b, but it will be exactly periodic, and the powerful theorems pertaining to periodic systems (Bloch's theorem, group theory, etc.) will apply, and the properties of the system that follow from symmetry may be calculated exactly. On the other hand, if the two lattice constants a and b are not commensurate, that is, their ratio a/b is equal to an irrational number, then the superposition is not periodic, and there is in general no easy way to calculate the properties. Nevertheless, the superposition does possess long-range order; it is constructed from simple rules, and its Fourier transform contains sharp peaks at π/a and π/b. The superposition of periodic lines with incommensurate lattice constants is an example of one-dimensional quasiperiodicity.

While quasiperiodicity in one dimension is fairly straightforward, in two and three dimensions the notion is considerably more interesting. For higher dimensions, another method of generating quasiperiodic patterns is used. One begins with a periodic lattice in a higher dimension (e.g., for a one-dimensional quasiperiodic lattice one uses a square lattice in two dimensions). The higher dimensional lattice is traversed with a lower dimensional surface (e.g., a one-dimensional straight line passing through the two-dimensional square lattice, making an angle θ with one of the lattice directions). Next a window width is defined, and lattice points falling within that window are projected onto the lower dimensional surface. If the direction cosines describing the orientation of the lower dimensional surface are irrational, then the sites projected onto the surface will form a quasiperiodic pattern. In the one-dimensional example a particularly interesting case occurs when the slope of the line is such that $\tan\theta = \tau = (\sqrt{5}+1)/2$, the golden mean (also known as the divine ratio), which is "the most irrational number." In this case the pattern of sites can be related to a Fibbonacci sequence, and rigorous theorems describing the properties of this special quasiperiodic symmetry can be derived. An algorithm that generates this sequence is

$$u_j = \tau - (\tau - 1)\left[\mathrm{Int}\left(\frac{j+1}{\tau}\right) - \mathrm{Int}\left(\frac{j}{\tau}\right)\right], \qquad (35)$$

where Int(x) indicates the integer part, and $j = 0, 1, 2, 3, \ldots$. The u_j will take on the values τ and 1 in a quasiperiodic sequence.

General theories for one-dimensional systems have been reviewed by Simon[9]; more recently renormalization group and dynamic mapping techniques have been used.[10] Some special properties of one-dimensional quasiperiodic systems are that the eigenvalue spectrum is a Cantor set, and the eigenfunctions may be extended, localized, or critical.

If one begins with a six-dimensional periodic lattice (e.g., a hypercubic lattice) and intersects it with a three-dimensional surface, then one may obtain a three-dimensional quasiperiodic pattern; such patterns may describe the recently discovered aluminum allow quasicrystals. If one begins with a five-dimensional periodic lattice and intersects it with a two-dimensional (plane) surface, then one may obtain a two-dimensional quasiperiodic pattern, or Penrose tilting[11, 12]; one such pattern is illustrated in Fig. 2. If one tries to tile a plane with only one tile shape (unit cell), then as in three dimensions only certain shapes and rotational symmetries are allowed. However, if one is allowed to use tiles of two or more shapes, then fivefold and other symmetries are possible. Algorithms for generating two- and three-dimensional Penrose tilings are quite complicated.[13, 14]

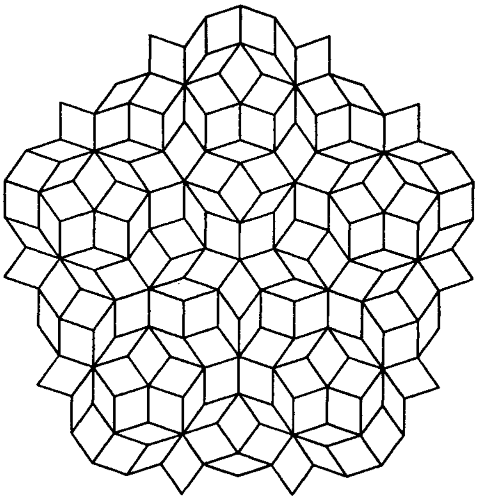


Fig. 2 Penrose tiling pattern consisting of two rhombuses whose areas are in the ratio of the golden mean, $(\sqrt{5}+1)/2$.

Although Penrose tile patterns are not periodic, they do have some symmetry properties.[11-15] Special patterns may have "inflation symmetry," so that a decoration of the tiles with certain lines will produce a replication of the original Penrose pattern, but with a reduced scale. There is also Conway's theorem, which states that given any local pattern (having some nominal diameter), an identical local pattern will be found within a distance of two diameters. The typical Penrose tiling shown in Fig. 2 is formed with two different rhombuses, a fat one and a skinny one; the ratio of the areas is the golden mean, $(\sqrt{5}+1)/2$.

Given a wave equation with a potential field having quasiperiodic symmetry, one can ask how the quasicrystalline symmetry affects the eigenvalues and eigenfunctions. Unlike the theorems for the quasiperiodic patterns in one dimension, theorems for two and three dimensions, if they exist, have not yet been discovered. For a pattern that is periodic, Bloch's theorem may be used to obtain exact solutions. For a pattern that is random, statistical methods may be used to make predictions about properties (e.g., Anderson localization). However a quasicrystalline pattern is not periodic, so that Bloch's cannot be used; and it is not random, so that statistical methods cannot be used. Experimental measurements[16] have been undertaken to determine the effects of quasicrystalline symmetry on eigenvalues and eigenfunctions, and it has been found that the eigenvalue spectrum occurs in bands and gaps whose widths are in the ratio of the golden mean. It will be interesting to see if a "Quasi-Bloch's" can be developed for two- and three-dimensional quasicrystals.

REFERENCES

1. A. A. Maradudin, E. W. Montroll, and G. H. Weiss, *Theory of Lattice Dynamics in the Harmonic Approximation*, Academic Press, New York, 1963.
2. J. M. Ziman, *Principles of the Theory of Solids*, Cambridge University Press, Cambridge, 1964.
3. J. E. Graebner, B. Golding, and L. C. Allen, "Phonon Localization in Glasses," *Phys. Rev.*, Vol. 34, 1986, pp. 5696–5701.
4. P. A. Lee and T. V. Ramakrishnan, "Disordered Electronic Systems," *Rev. Mod. Phys.*, Vol. 57, 1985, pp. 287–337.
5. B. L. Al'Tshuler and P. A. Lee, "Disordered Electronic Systems," *Phys. Today*, Vol. 41, December, 1988, pp. 36–44.
6. M. Luban and J. H. Luscombe, "Localized Eigenstates of One Dimensional Tight-binding Systems: A New Algorithm," *Phys. Rev.*, Vol. B35, 1987, pp. 9045–9055.
7. H. Furstenberg, "Noncommuting Random Products," *Trans. Am. Math Soc.*, Vol. 108, 1963, pp. 337–428.
8. D. Shechtman, I. Blech, D. Gratias, and J. W. Cahn, "Metallic Phase with Long Range Orientational Order and No Translational Symmetry," *Phys. Rev. Lett.*, Vol. 53, 1984, pp. 1951–1953.
9. B. Simon, "Almost Periodic Schrodinger Operators: A Review," *Adv. Appl. Math.*, Vol. 3, 1982, pp. 463–490.
10. M. Kohmoto and J. R. Banavar, "Quasiperiodic Lattice: Electronic Properties, Phonon Properties and Diffusion," *Phys. Rev.*, Vol. B34, 1986, pp. 563–566.
11. R. Penrose, *Bull. Instit. Math. Appl.*, Vol. 10, 1974, p. 266.
12. M. Gardner, "Mathematical Games: Extraordinary Nonperiodic Tiling That Enriches the Theory of Tiles," *Sci. Am.*, Vol. 236, 1977, pp. 110–115.
13. D. Levine and P. J. Steinhardt, "Quasicrystals. I. Definition and Structures," *Phys. Rev.*, Vol. B34, 1986, pp. 596–616.
14. J. E. S. Socolar and P. J. Steinhardt, "Quasicrystals. II. Unit-Cell Configurations," *Phys. Rev.*, Vol. B34, 1986, pp. 617–647.
15. P. J. Steinhardt, "Quasicrystals," *Am. Sci.*, Vol. 74, 1986, pp. 586–597.
16. S. He and J. D. Maynard, "Eigenvalue Spectrum, Density of States, and Eigenfunctions in a Two-dimensional Quasicrystal," *Phys. Rev. Lett.*, Vol. 62, 1989, pp. 1888–1891.

58

SURFACE WAVES IN SOLIDS AND ULTRASONIC PROPERTIES

MOISÉS LEVY AND SUSAN C. SCHNEIDER

1 INTRODUCTION

Surface acoustic wave (SAW) devices are used for many applications, particularly for high-frequency electronic signal processing. In order to optimize the SAW devices, the properties of the piezoelectric substrates, SAW interdigital electrode (IDT) optimization techniques, and SAW temperature compensation techniques have all been extensively studied.[1–4] When a SAW device is used in experiments to study thin films, however, the SAW device is used as both the vehicle to launch the SAW waves and as a support for the thin film. In such experiments, the SAW velocity and attenuation measured are determined by *both* the thin film and the SAW device. In this chapter, a brief review of the fundamentals of SAW propagation through metallic layers deposited on substrates and the SAW velocity of the layered system is given in Section 3. The attenuation of SAWs due to electron–phonon interaction in both the normal and superconducting state is described in Section 4.1; the attenuation of SAWs due to the acousto-electric effect in both normal and superconducting state for homogeneous and granular metallic films is discussed in Section 4.2. The quantum Hall effect as it is related to SAW attenuation is summarized in Section 4.3, and the attenuation of SAWs due to magnetoelastic interaction is described in Section 4.4.

2 SAW DEVICES

A drawing to represent a "typical" SAW device used in SAW attenuation experiments is shown in Fig. 1. The SAW device consists of a piezoelectrically active substrate, such as quartz or $LiNbO_3$, on which one or more sets of IDTs are deposited. In some cases, the SAW device is fabricated with a piezoelectrically active coating such as ZnO or AlN deposited on a base substrate material such as fused quartz or sapphire. The IDTs are fabricated with a microphotolithographic process. The fundamental frequency of operation of the SAW device is determined by the width of individual electrodes as well as the spacing between the electrodes and the SAW velocity of the substrate. In addition, the design of the IDTs can be optimized for the generation of overtone frequencies to allow the frequency dependence of an attenuation mechanism to be studied.

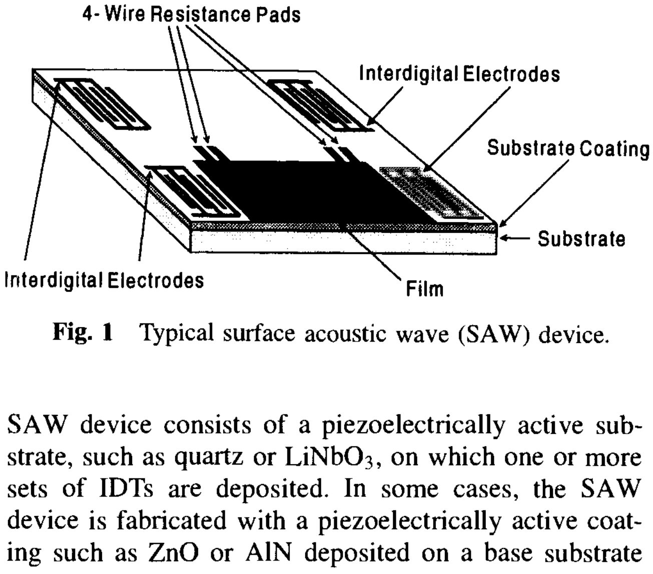


Fig. 1 Typical surface acoustic wave (SAW) device.

When a SAW device is used in experiments to study the properties of thin films, the material to be studied may be deposited on the substrate between one of the sets of IDTs, as shown in the figure. When available on the SAW device, a second set of IDTs is used to monitor the attenuation of a "clear path" on which no film has been deposited. The measurements taken on the clear path provide information about the background attenuation of the substrate. To simultaneously monitor the resistance of a

---

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker
Copyright © 1997 John Wiley & Sons, Inc.

thin film using four-wire resistance measurements, many devices have additional electrodes deposited directly on the film (or on the substrate) that do not interfere with the sound path in the film.

3 SURFACE ACOUSTIC WAVE VELOCITY IN FILM–SUBSTRATE GEOMETRIES

When a thin film is deposited on the surface of a SAW device, the SAW velocity of the film–substrate combination is determined by the elastic constants, density, and thickness of the thin film as well as the elastic constants and density of the (piezoelectric) substrate. In general, the SAW velocity for the film–substrate combination can be obtained by solving the equations of motion in the film and the substrate subject to the appropriate boundary conditions at the free surface of the film and the interface between the film and substrate.[5] The simplest case for this calculation is when both the substrate and the film are considered to be isotropic elastic bodies. The calculation method for the SAW velocity for this case is outlined here.

In the following it is assumed that the SAW with wave vector $\boldsymbol{q}$ propagates along the x-axis, and the z-axis is perpendicular to the film plane. The crests of the plane wave are along the y-axis. The film thickness is h. The film substrate interface is the $z = 0$ plane and the free surface of the film is the plane is $z = h$; the substrate lies in the region where $z < 0$. The surface wave is assumed to have the general form $\boldsymbol{\psi}(\boldsymbol{r}, t) = (\alpha, \beta, \gamma) e^{bqz} e^{i(qx - \omega t)}$, where α, β, and γ are the partial-wave coefficients of the wave and ω is the angular frequency of the wave. In terms of the displacement vector $\boldsymbol{\psi}$, the equation of motion has the form

$$\rho \ddot{\boldsymbol{\psi}} = \boldsymbol{\nabla} \cdot \boldsymbol{T}, \tag{1}$$

where the stress tensor $\boldsymbol{T}$ is defined as

$$T_{ij} = \lambda \sum_k \epsilon_{kk} \delta_{ij} + 2\mu \epsilon_{ij}. \tag{2}$$

The strain is defined by $\epsilon_{ij} \equiv (\nabla_i \psi_j + \nabla_j \psi_i)/2; \lambda$ and μ are the isotropic Lamé constants, and ρ is the density of the material. Equation (1) may be rewritten in terms of the displacement to give the equation of motion

$$\rho \ddot{\boldsymbol{\psi}} = (\lambda + \mu) \boldsymbol{\nabla} (\boldsymbol{\nabla} \cdot \boldsymbol{\psi}) + \mu \boldsymbol{\nabla}^2 \boldsymbol{\psi}. \tag{3}$$

In order to find $\boldsymbol{\psi}$ for the layered half-space, the equation of motion is solved separately for each medium. The resulting solutions must then satisfy the three boundary conditions, which may be summarized as follows:

1. Stress free surface: $T^F_{zz} = T^F_{zx} = 0$ at $z = h$.
2. Continuous stress: $T^F_{zz} = T^S_{zz}$ and $T^F_{zx} = T^S_{zx}$ at $z = 0$.
3. Continuous amplitude: $\psi^F = \psi^S$ at $z = 0$.

The superscripts F and S refer to the film and substrate, respectively.

The resulting six equations from the above boundary conditions can be written in matrix form as

$$\begin{pmatrix} a & 1 & -a_F & -1 & -a_F & -1 \\ -1 & -d & 1 & d_F & -1 & -d_F \\ A\dfrac{\mu}{\mu_F} & 2d\dfrac{\mu}{\mu_F} & -A_F & -2d_F & A_F & 2d_F \\ -2a\dfrac{\mu}{\mu_F} & -A\dfrac{\mu}{\mu_F} & 2a_F & A_F & 2a_F & A_F \\ 0 & 0 & A_F e^{a_F qh} & 2d_F e^{d_F qh} & -A_F e^{-a_F qh} & -2d_F e^{-d_F qh} \\ 0 & 0 & -2a_F e^{a_F qh} & -A_F e^{d_F qh} & -2a_F e^{-a_F qh} & -A_F e^{-d_F qh} \end{pmatrix} \cdot \begin{pmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \\ C_6 \end{pmatrix} = 0, \tag{4}$$

where the subscript F refers to the film. The coefficients without subscripts refer to the substrate.

For the substrate

$$a = \left(1 - \frac{v^2}{v_t^2}\right)^{1/2}, \qquad d = \left(1 - \frac{v^2}{v_l^2}\right)^{1/2}, \tag{5}$$

where $v_l = [(\lambda + 2\mu)/\rho]^{1/2}$ and $v_t = (\mu/\rho)^{1/2}$ are the bulk longitudinal and transverse velocities of the substrate, v is the SAW velocity, and $A = 1 + a^2$.

For the film

$$a_F = \left(1 - \frac{v^2}{v_{tF}^2}\right)^{1/2}, \qquad d_F = \left(1 - \frac{v^2}{v_{lF}^2}\right)^{1/2}, \tag{6}$$

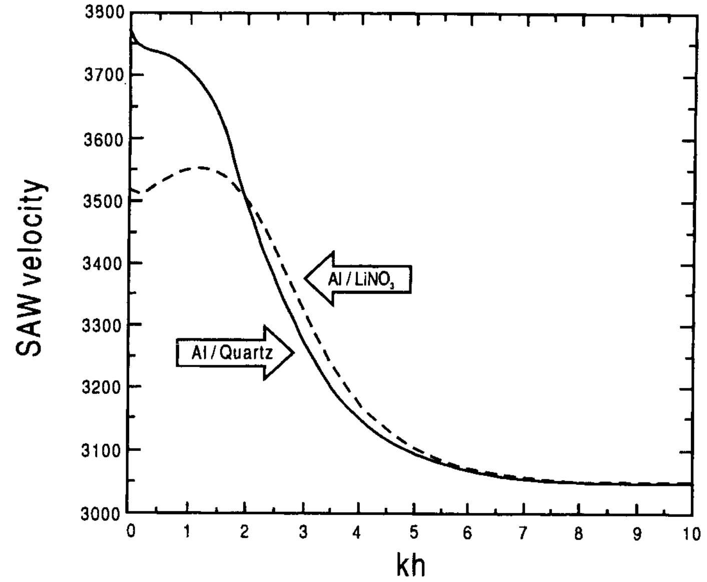


Fig. 2 SAW velocity calculation as a function of *qh* for an aluminum film on a quartz substrate (solid line) and a $LiNbO_3$ substrate (dashed line).

where $v_{lF} = [(\lambda_F + 2\mu_F)/\rho_F]^{1/2}$ and $v_{tF} = (\mu_F/\rho_F)^{1/2}$ are the longitudinal and transverse bulk velocities of the film and $A_F = 1 + a_F^2$. Equation (4) has a nontrivial solution only if the determinant of the coefficient matrix equals zero. The velocity of the surface wave is the only undetermined parameter in this determinant. A computer program is used to search for those values of the velocity that make the determinant vanish for a chosen frequency ω and particular value of h. The results of the SAW velocity calculation for an aluminum film on a $LiNbO_3$ substrate and a quartz substrate are shown in Fig. 2. The velocity of the SAW in the film–substrate combination is determined by the substrate Rayleigh velocity at zero film thickness, and as the film thickness increases, the SAW velocity approaches the Rayleigh velocity of the film material.

4 SAW ATTENUATION IN FILM–SUBSTRATE GEOMETRIES

Just as the presence of a film on a substrate results in a change in the SAW velocity, the presence of the film can also lead to a change in the attenuation of the SAW. The SAW devices are not lossless, but in general, the attenuation of the SAW due to the substrate alone is known or can be determined. Any additional attenuation of the SAW measured when a film has been deposited on the substrate can then be attributed directly to the presence of the film. The SAW attenuation mechanisms in layered film–substrate combinations can be placed into two broad classes. Surface acoustic wave attenuation produced by the electron–phonon interaction in a metallic film is an example of the "mechanical class" of SAW attenuation mechanisms. In the case of electron–phonon interaction, it is the film that is primarily responsible for the observed SAW attenuation, with the substrate serving to launch the SAW, support the film, and carry the mechanical disturbance. The other class of SAW attenuation mechanisms in layered film–substrate combinations can be characterized as "electrical" in nature, such as SAW attenuation produced through the acoustoelectric effect. In the acoustoelectric effect, a metallic film interacts with the piezoelectric substrate by electrically "shorting out" the surface of the substrate; in this case, the substrate launches the SAW and carries the electrical disturbance but does not necessarily serve as a support for the film. Regardless of the class of mechanism involved in producing the SAW attenuation in the film–substrate calculations, the SAW attenuation is proportional to the energy absorbed by the SAW in the material divided by the total mechanical energy of the SAW in both the film and substrate.

4.1 Attenuation of SAW due to Electron–Phonon Interaction

The first measurement of the attenuation of bulk waves due to electron–phonon interaction was reported by Bommel on a single crystal of lead in 1954.[6] Theoretical models to describe the attenuation of longitudinal and transverse bulk waves produced by electron–phonon interaction include models developed by Mason,[7] Morse,[8] and Pippard,[9] for example. The

SAW measurements in which the attenuation is due to electron–phonon interaction have been performed on a wide variety of films.[10–18] At present, there are no ab initio calculations for the electron–phonon interaction in thin films as there are for bulk waves. However, there are several models[19,20] for SAW attenuation that incorporate the results of the theoretical models developed for bulk wave attenuation produced by electron–phonon interaction. For this reason, the attenuation of sound waves in bulk materials due to electron–phonon interaction in both the normal and superconducting states will be summarized.

Electron–Phonon Interaction in Bulk Materials

Pippard[9] found that the relevant parameter for predicting the attenuation of bulk sound waves in the normal state for a metal is ql, where q is the phonon wave vector and l is the electron mean free path. When $\omega\tau \ll 1$ and $ql \ll 1$, where $\omega = qv_l$ is the angular frequency and $\tau = l/v_F$, the attenuation of longitudinal waves is given by

$$\alpha_L = \frac{4}{15}\,\frac{Nmv_F^2\omega^2\tau}{\rho v_l^3}. \tag{7}$$

The attenuation of transverse sound waves due to electron–phonon interaction in the same limit is given by

$$\alpha_T = \frac{1}{5}\,\frac{Nmv_F^2\omega^2\tau}{\rho v_t^3}. \tag{8}$$

In these equations, N is the number of electrons per unit volume, v_F is the Fermi velocity, v_l and v_t are the materials's longitudinal and transverse sound velocities, respectively, ρ is the mass density of the material, and τ is the relaxation time for electron–phonon interaction. Note that for both cases the attenuation is linearly dependent on the relaxation time and quadratically dependent on the frequency. Many other authors have derived similar expressions for the attenuation.[7,8,21–23]

In the superconducting state, the ultrasonic attenuation of bulk waves is given by the Bardeen, Cooper, and Schrieffer[24] (BCS) expression

$$\frac{\alpha_s}{\alpha_n} = \frac{2}{e^{\Delta/kT} + 1}. \tag{9}$$

This result for the ratio of the attenuation coefficient in the superconducting state, α_s, to that in the normal state, α_n, was derived in 1957[24] for the ultrasonic attenuation coefficient for longitudinal waves in the limit $ql \gg 1$. Subsequently, it was shown that this result is also valid for both longitudinal and transverse waves in the limit $ql \ll 1$,[25] the limit in which the SAW attenuation data are generally acquired.

Electron–Phonon Interaction in Thin Films

The computation of the attenuation due to electron–phonon interaction produced by a thin metallic film deposited on a piezoelectric substrate is complicated by several factors. These factors are (a) the attenuation is only produced in the film; (b) the substrate has different elastic constants and density than the film; (c) the longitudinal SAW motion is *not purely compressive* and the transverse SAW motion is *not purely shear*; and (d) theoretical calculations are performed in the limit where the piezoelectric interaction is considered to be negligible.

In earlier derivations it was assumed that either the displacement[19] or the strains[26] were continuous. A later approach[20] satisfied both of these boundary conditions. This later approach to calculating the attenuation of SAWs due to electron–phonon interaction combines one of the techniques used to theoretically model the attenuation of bulk waves due to electron–phonon interaction[7] with the method described previously to calculate the velocity of SAW propagation in film–substrate geometries. The attenuation of SAW is calculated by including viscous loss terms in the macroscopic equation of motion. The viscous loss terms were proposed by Mason[7] to calculate the attenuation of bulk sound waves due to electron–phonon interaction in the limit $ql \ll 1$. In terms of the displacement vector $\boldsymbol{\psi}$ the equation of motion is given by Eq. (1), where the stress tensor $\boldsymbol{T}$ includes viscous damping terms and is defined as

$$T_{ij} = \lambda \sum_k \epsilon_{kk}\delta_{ij} + 2\mu\epsilon_{ij} + \chi \sum_k \dot{\epsilon}_{kk}\delta_{ij} + 2\eta\dot{\epsilon}_{ij}. \tag{10}$$

As before, the strain is defined by $\epsilon_{ij} \equiv \frac{1}{2}(\nabla_i\psi_j + \nabla_j\psi_i)$; λ and μ are the isotropic Lamé constants. The compressional and shear viscosity coefficients χ and η are now explicitly included in the equation of motion. The bulk modulus is related to λ and μ by $B = \lambda + 2\mu/3$; and one may relate the bulk coefficient of viscosity κ to the compressional viscosity χ and shear viscosity η through a similar equation, $\kappa = \chi + 2\eta/3$. The bulk coefficient of visocity is equal to zero for a monatomic gas, such as a free electron gas, since uniform compressions in such a system would produce no viscous losses; and therefore, $\chi = -2\eta/3$. The shear viscosity coefficient η was originally derived by Mason[7] in his calculation of the attenuation of bulk waves due to electron–phonon interaction, $\eta = Nmv_F^2\tau/5$. When these definitions of the shear and compressional viscosity, together with the time dependence of the SAW, $\boldsymbol{\psi}(\boldsymbol{r}, t) = \boldsymbol{\psi}(\boldsymbol{r})e^{-i\omega t}$, are used, the following equation of motion for the SAW results:

$$-\rho\omega^2\boldsymbol{\psi}(\boldsymbol{r}) = (\lambda + \mu - \tfrac{1}{3}i\omega\eta)\nabla(\nabla \cdot \boldsymbol{\psi}(\boldsymbol{r})) + (\mu - i\omega\eta)\nabla^2\boldsymbol{\psi}(\boldsymbol{r}). \tag{11}$$

This equation is the same as that of a nonlossive medium where λ is replaced by $\lambda + \frac{2}{3}i\omega\eta$ and μ by $\mu - i\omega\eta$. Hence, the solutions for the displacement fields and the propagation velocity are the analytic continuation of the solutions to the equation of motion without explicit losses and can be found by an extension of the same methods described previously to calculate the velocity of SAW in layered film–substrate geometries when the longitudinal and transverse velocities of the film are modified to take into account the explicit loss produced by the viscous damping; $v_{lF} = [(\lambda_F + 2\mu_F - 4i\omega\eta/3)/\rho_F]^{1/2}$ and $v_{tF} = [(\mu_F - i\omega\eta)/\rho_F]^{1/2}$ are the longitudinal and transverse bulk velocities of the film. Thus the viscous terms make the coefficients in the matrix complex. The same procedure described previously to determine the SAW velocity is followed to determine the SAW attenuation; that is, the matrix has a nontrival solution only if the determinant of the coefficient matrix equals zero. The velocity of the surface wave is the only undetermined parameter in this determinant, but this time, the velocity is a complex quantity. The amplitude attenuation is then given by the imaginary part of the complex wave vector q, $\alpha_{\mathrm{amp}} = \mathrm{Im}(q) = \mathrm{Im}(\omega/v)$. The amplitude attenuation of the SAW as a function of qh for an aluminum film on either a quartz or a LiNbO3 substrate is shown in Fig. 3. The value of η for aluminum used in this calculation was $\eta = 2.19 \times 10^{-3}$ P. In the limit of $qh \ll 1$, the SAW attenuation calculated with this approach appears to be linear with respect to qh.

The derivation for the attenuation of SAW due to electron–phonon interaction in the superconducting state also exploits the results obtained for the attenuation of bulk waves due to electron–phonon interaction in bulk materials. Since the attenuation of a SAW can be ascribed to the shear viscosity of the electron gas, and since the shear viscosity of the electron gas follows the BCS region as derived for transverse bulk waves, which are purely shear, it should follow that the attenuation of a surface acoustic wave in the $ql \ll 1$ limit should follow the BCS relation

$$\left(\frac{\alpha_s}{\alpha_n}\right)_{\mathrm{SAW}} = \frac{2}{e^{\Delta/kT} + 1}. \tag{12}$$

This same result is also obtained from models that ascribe the attenuation of the SAW to the sum of the compressive and shear losses, since both of these losses follow the BCS relation in the $ql \ll 1$ limit.

Magnetic Field Effects. Superconductors may be divided into two types, types I and II. This classification depends upon whether the surface energy required to produce a superconducting–normal state interface is positive or negative, which in turn is determined by the ratio of the London penetration length λ to the coherence length ξ.

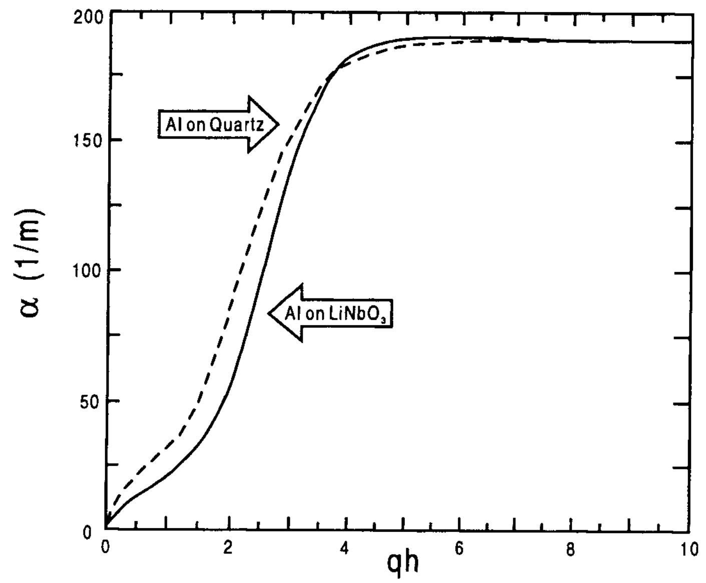


Fig. 3 Amplitude attenuation of the SAW as a function of qh for an aluminum film on either a quartz or a LiNbO3 substrate.

A type I superconductor has $\lambda/\xi < 1$, since the superconducting condensation energy expended for producing a normal cylinder of radius ξ in a superconductor will be higher than the energy gained by allowing a magnetic field to occupy a cylinder of radius λ around the same axis. A type II superconductor has $\lambda/\xi > 1$ for the opposite argument. Therefore, a type I superconductor is perfectly diamagnetic up to the thermodynamic critical field H_c, whereupon the magnetic field totally penetrates the superconductor, destroying superconductivity. Sound waves traveling through the interior of the superconductor do not sense the presence of an external magnetic field and therefore are not affected; thus the attenuation remains constant up to H_c and regains the normal state value when superconductivity is quenched. A type II superconductor remains in the diamagnetic state up to a lower critical field H_{c1}, whereupon magnetic flux lines of value $\phi = h/2e$ and radius ξ start to penetrate (h is Planck's constant and e is the electronic charge). The superconductor remains in this vortex state, in which normal-state cylinders coexist in a superconducting matrix up to the upper critical field H_{c2} when the flux lines coalesce. The magnetization immediately below H_{c2} is linearly dependent upon the applied magnetic field. Type II superconductors should be further divided into dirty and clean type II behavior. In dirty type II superconductors, the electron mean free path is shorter than the coherence length. Electrons inside one of the normal cylinders suffer several collisions before escaping and therefore contribute the normal-state value to the attenuation. Since the fraction of volume occupied by flux lines is proportional to the quantity 1 minus the magnetization $(1-M)$, the attenuation should depend linearly[27] upon the applied magnetic field, $(\alpha_s - \alpha_n)/\alpha_n \simeq (H_{c2} - H)$. In clean type II superconductors, the electron mean free path is longer than the coherence length, and the electrons take a spatial average of the superconducting order parameter Δ, which is proportional to the square root of the magnetization. By placing such an average for Δ in the BCS expression for the attenuation [Eq. (12)] and making a small value expansion, it is found[28] that $(\alpha_s - \alpha_n)/\alpha_n \approx \sqrt{H_{c2} - H}$. Thus for dirty type II superconductors, the attenuation should depend linearly on the applied magnetic field, while for clean type II superconductors, it should depend upon the square root of the difference between the upper critical field H_{c2} and the applied field H. An example of the linear dependence may be found in measurements of Fredricksen et al.[29] when SAWs were used to measure the attenuation coefficient of a Nb_3Sn film near H_{c2} at 700 MHz.

Experimental Measurements of SAW Attenuation due to Electron–Phonon Interaction In 1969, Akao[10] and Krätzig et al.[11] reported the first SAW measurements done on superconducting films. Akao[10] measured attenuation as a function of temperature on In and Pb films. Krätzig et al.[11] reported the magnetic field dependence of the SAW attenuation in a Pb film with the magnetic field either parallel or perpendicular to the film surface. In 1974, Robinson et al.[12] reported on the temperature dependence of SAW attenuation in a granular Al film. In 1978, Park et al.[13] reported on SAW attenuation measurements as a function of temperature in PbCu and PbAg sandwiches. In 1979, Barley and Marshall[15] reported attenuation measurements as a function of temperature in Zn films. That same year, Fredricksen et al.[14] reported on magnetic field measurements on Nb_3Sn. In 1980, Salvo et al.[16] reported on SAW attenuation measurements as a function of temperature on Nb_3Ge and Fredricksen et al.[17] reported similar measurements on Nb_3Sn. In 1991, Baum et al.[18] reported attenuation measurements for a thin film of niobium. Many of these films appear to follow what could be called the classical BCS relation [Eq. (9) and (12)] exactly. The other films, such as Nb_3Ge, exhibit deviations from this relation. The results of several of these SAW attenuation measurements in superconducting thin films are summarized in Table 1. The parameters reported in this table include the superconducting transition temperature T_c reported for the film and the zero-temperature energy gap $2\Delta(0)$. Shown in Fig. 4 is an example of SAW attenuation data for a Nb_3Sn film on a $LiNbO_3$ substrate when compared to BCS theory [Eq. (12)]. The zero-temperature energy gap for this film is $2\Delta(0) = 3.5kT_c$.

In addition to determining the zero-temperature energy gap and the superconducting critical temperature, SAW attenuation measurements can also be used to determine a value for the electron mean free path in the material. Using the value of the normal-state attenuation obtained from experimental measurements, a value for the electron mean free path can be obtained, $l_{\alpha,\mathrm{SAW}}$, which can be compared to the value of the electron mean free path obtained from conductivity/resistivity measurements on the films, l_{elec}. The value for $l_{\alpha,\mathrm{SAW}}$ may be deduced from plots such as those shown in Fig. 3 by selecting the appropriate value for τ in η. An alternate and simpler approach, though approximate, would be to use the procedure outlined in Ref. 19. Values for the electron mean free path obtained from both SAW attenuation measurements and resistivity measurements are shown in Table 2 for three superconducting films that are granular in nature. In all cases shown in Table 2 the value of the mean free path obtained from SAW attenuation measurements is greater than that obtained from the electrical resistivity measurement. In the SAW attenuation measurements, the value of the mean free path obtained is consistent with one given by diffuse scattering over the surface of grains in the film. The SAWs appear to sample the granules in the film in parallel and

TABLE 1 Superconducting Energy Gap as Determined by SAW Attenuation Measurements on Thin Films

Film	Thickness, h (Å)	T_c (K)	$2\Delta(0)/kT_c$	Reference
Nb_3Sn	5000	18	3.5	17
Aluminum	300	1.7	3.6	12
Zinc	3000	1.5	4.2	15
Zinc	10,000	1.31	3.8	15
Indium	500, 1500	—	4.3	10
Lead	1000, 2500	—	3.9	10
Niobium	3000	8.5	3.5	18
Nb_3Ge	5000	21	3.5	16

are thus sensitive to the longest mean free path in the granules. In the resistivity measurements for l, the electrical current passes through the granules in series, and therefore, resistivity measurements are sensitive to the smallest mean free path in the granules. Thus, the two measurements for electron mean free path yield complementary information.[19]

4.2 Acoustoelectric Effect

The attenuation of SAW in piezoelectrics coated with semiconducting films has been described by Ingebrigtsen[30] and Adler,[31] for example. Adler used an intuitive approach to derive an expression for the attenuation (gain or loss) of an acoustic surface wave that was related to frequency, the effective electromechanical coupling constant K, and the semiconductor drift velocity. This same type of intuitive approach can be used to describe the acoustoelectric interaction of SAWs with superconducting films.

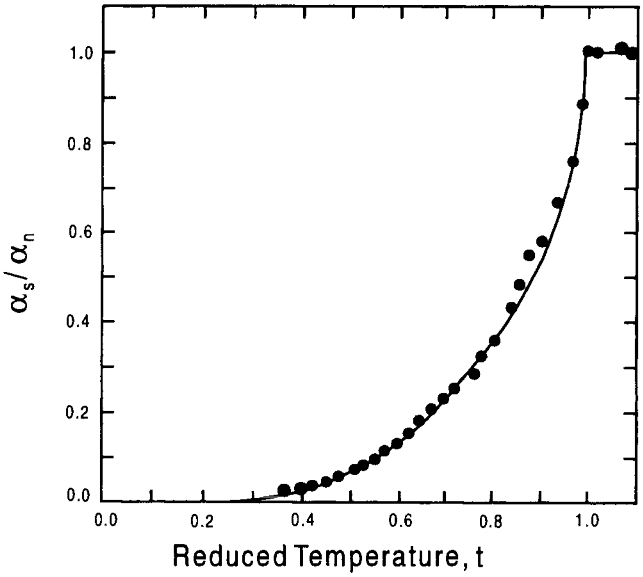


Fig. 4 SAW attenuation data as a function of reduced temperature, $t = T/T_c$, for a Nb_3Sn film on a $LiNbO_3$ substrate when compared to BCS theory [Eq. (9)].

The effect of acoustoelectric interaction with superconducting films and with a two-dimensional electron gas that is in close proximity to a piezoelectric surface carrying a SAW may be described as follows. A SAW propagating through a piezoelectric substrate produces surface electric polarization charges associated with the SAW strains. When a conducting film either is deposited on or is brought into close proximity to the piezoelectric substrate surface, the polarization charges in the substrate induce image charges in the conducting film. These image charges will move in the conducting film with the velocity of the SAW, producing energy dissipation that attenuates the SAW. If the conductivity of the film is infinite, there will be zero energy dissipation; if the conductivity is zero, there will be no image charges and the dissipation will again be zero. Only if the conductivity is finite will there be energy dissipation. Therefore, the attenuation process may be described by a debye relaxation expression, which for the case when the film is deposited directly onto the piezoelectric substrate is given by

$$\alpha = \frac{\kappa^2}{v_0} \frac{\omega^2 \tau}{1 + \omega^2 \tau^2}, \tag{13}$$

where α is the energy attenuation coefficient in nepers per centimetre, κ^2 is the electromechanical coupling con-

TABLE 2 Electron Mean Free Path Comparison

Film	$l_{\alpha,\text{SAW}}$ (Å)	l_{elec} (Å)	Reference
Aluminum	33	1.3	19
Niobium	1050	450	18
Nb_3Ge	140	10	16

stant, v_0 is the velocity of the SAW, ω is the angular frequency of the SAW, and τ is a relaxation time. For very thin films, τ is given by $\omega\tau = v_0(\epsilon_0 + \epsilon)/\sigma_\square$, where ϵ_0 and ϵ are the dielectric permittivity of vacuum and the substrate, respectively, and $\sigma_\square$ is the sheet conductivity of the conducting film ($\sigma_\square = \sigma d$, where d is the thickness of the film and σ is the electrical conductivity of the film material). Very thin films are defined as having a thickness that is smaller than $\lambda/2$, where λ is the wavelength of the SAW. The attenuation α will have a maximum when $\omega\tau = 1$, which occurs for $\sigma_{2D} = v_0(\epsilon_0 + \epsilon)$, generally equal to about $10^{-6}\,\Omega^{-1}$; σ_{2D} is known as the characteristic conductivity of the system.

Since Eq. (13) is a typical relaxation expression for the attenuation, the SAW velocity will also change with $\omega\tau$ as follows[31]:

$$\frac{\Delta v}{v_0} = -\frac{\kappa^2}{2}\,\frac{1}{1+\omega^2\tau^2}, \tag{14}$$

so that, for $\omega\tau \rightarrow \infty$, $\Delta v/v \rightarrow 0$, and for $\omega\tau \rightarrow 0$, $\Delta v/v_0 = -\kappa^2/2$. This behavior should be expected since the case $\omega\tau \rightarrow \infty$ results when the piezoelectric surface is exposed to free space and then the normal SAW velocity should be obtained. The latter case, $\omega\tau \rightarrow 0$, results when the surface is covered by a perfectly conducting film. This perfectly conducting film shorts out the surface electric fields, thus quenching the electrical energy carried by the wave and leaving only the mechanical energy; since the elastic constants are given by the second partial derivative of the Gibbs free energy, they should also be reduced.

It is convenient to recast Eq. (13) in terms of σ_{2D} as

$$\alpha = \frac{\omega\kappa^2}{v_0}\,\frac{\sigma_\square/\sigma_{2D}}{1+(\sigma_\square/\sigma_{2D})^2}. \tag{15}$$

For $\sigma_\square \ll \sigma_{2D}$, the joule loss in the film is of the form of a constant voltage source, V^2/R, since the acoustoelectric currents through the film are very small. However, for $\sigma \gg \sigma_{2D}$, there is enough redistribution of charge in the conductor to almost completely screen the electric field and the loss is of the form of I^2R, such as for a constant current generator.

In some cases, it is not possible to deposit the film onto a piezoelectric substrate directly. Instead, the film is deposited on another substrate, which is then brought into close proximity to the piezoelectric substrate. Schenstrom[32] has derived an expression for proximity coupling between the surface of the SAW substrate and the conducting film.

Even superconducting films with the highest normal-state resistivities have values of $\sigma_\square$ that are much larger than σ_{2D}. Therefore, in such films, $\omega\tau < 1$, and it is possible to simplify Eq. (13) to

$$\alpha = \frac{\kappa^2}{v_0}\,\omega^2\tau = \frac{\omega\kappa^2}{v_0}\,(\epsilon_0 + \epsilon)R_\square, \tag{16}$$

where $R_\square = 1/\sigma_\square$. Equation (16) implies that in the limit of $\omega\tau < 1$ the attenuation should be proportional to the sheet resistivity of the conducting film. If the conducting film undergoes a superconducting transition, then the SAW attenuation produced by the acoustoelectric interaction should vanish. Any change in attenuation should be proportional to the change in the resistivity of the film, since the dielectric constants and the electromechanical coupling constants should not be affected by the superconducting transition. Schmidt[33] has shown that the change in attenuation produced by superconducting quenching of the acoustoelectric effect is proportional to the change of resistance in an amorphous indium film. The broadness of the transition in this film may be accounted for by superconducting fluctuations[34] above the midpoint of the transition and by the motion of unbound vortex pairs below the midpoint.[35] This justifies the deduction that the attenuation due to the acoustoelectric effect should be proportional to the sheet resistivity through the whole of a superconducting transition.

However, data obtained on a granular lead film[36] do not exhibit this proportionality in the superconducting state. Not only is the acoustoelectric attenuation larger than would be predicted by Eq. (16), it also has a finite value in the superconducting state even when the sheet resistivity has vanished. To account for both the shape of the attenuation curve in the superconducting state and its magnitude with respect to the sheet resistivity curve, a percolation model was developed. The percolation model for attenuation produced by the acoustoelectric effect is motivated by the granular nature of some of the films studied. It is assumed that the grains in the film are separated by oxide barrier junctions through which electrical currents may tunnel and, for simplicity, that these tunneling junctions are arranged in a two-dimensional square lattice network. The essence of the model is that measurements of DC electrical resistivity sample the percolation resistivity of such a network with infinite dimensions, or at least the dimensions of the film, while the acoustoelectric attenuation coefficient depends on the average resistivity of squares whose dimensions are comparable to the SAW wavelength,[37] specifically $\lambda/2\pi$.

The percolation resistance of such networks can be determined using a model due to Ambegaokar, Halperin, and Langer (AHL).[38] The tunneling junctions are replaced by resistors whose values correspond to those of the junctions. All these resistors are removed from the network and then returned to the network according to the following scheme: First, the lowest value resistors are returned at random, then the next lowest value resistors are returned, and so on. When the fraction of resistors that have been returned to the network is equal to the critical percolation fraction—$\frac{1}{2}$ for the case of an infinite square network—a percolation path will be established through the resistor. For a two-dimensional network, the sheet resistivity R_n is approximately equal to that value of the critical resistor r_c, which establishes this percolation path. This claim is plausible, since resistors with values larger than r_c do not contribute much to the sheet resistivity because they are shunted by resistors whose values are on the order of r_c; resistors with values less than r_c cannot by themselves provide current transport over macroscopic distances. The assumptions made by Levy et al.[37] yield a parabolic distribution function for resistors with a minimum resistance r_1 and a maximum resistance r_2, and $R_n = (r_1 r_2)^{1/2}$. As mentioned above, R_n is a measure of the sheet resistivity of a film with a DC potential or an AC potential with infinite wavelength. The acoustoelectric effect produces AC potentials with finite wavelengths, namely the SAW wavelength λ. Therefore, the AC potentials sample the resistance of small squares whose dimensions are comparable to λ. In fact, Levy et al.[37] postulate that this dimension should be $l = \lambda/2\pi$. It is easy to deduce the factor of 2 in this expression since the electric potential changes sign twice over a wavelength. The factor of 2π comes from the belief that since most acoustic effects have some sort of transition when $ql \simeq 1$, then ql would be the important parameter in any acoustic propagation equation. This dimension l can be conveniently expressed in terms of the average number of grains L contained in l. Since the granular film is inhomogeneous, the local sheet resistance of these small sections, $R(L)$, will vary from location to location. Therefore, the ultrasonic attenuation in Eq. (16) should be governed by the average sheet resistance $\boldsymbol{R}(L)$ of these small sections of the film, each of which can be represented by a finite resistor network, and not the sheet resistance of the entire film. It was found[33] that films with over 300 grains per $\lambda/2\pi$ may be considered homogeneous for the purpose of acoustoelectric attenuation measurements.

As the temperature is lowered, the metallic grains become superconducting at their transition temperature T_g, but the Josephson junctions between the grains remain resistive until a temperature is reached where the Josephson coupling energy E_j is equal to or higher than the thermal energy. Thus, as the temperature is lowered, more junctions become superconducting by passing Josephson currents, the lower resistance junctions becoming superconducting first. The percolation resistance of the film is then computed using an effective medium approximation[39] wherein the junction resistances have a binary distribution; the junctions are either superconducting with a zero resistance if the temperature is below their particular E_j or the junctions have a resistance value equal to R_n. Then the sheet resisivity of the film may be calculated by assuming a linear dependence of the overall resistance with respect to the fraction of R_n resistors that are present, the resistance going to zero when half of the junctions become superconducting. This binary distribution is adequate for describing the resistivity of the small sections sampled with SAW since there are only a few tunneling junctions per wavelength. However, for computing the DC resistivity of films that encompass a very large number of junctions per wavelength, it is necessary to take into account the full distribution of the junction resistances in the normal state.[40] By doing so, even when plotting the resistivity data logarithmically, Schmidt et al.[40] find that this approach fits the resistivity data for several granular films very well.

The same procedure may be used to obtain the sheet resistance of each individual finite resistor network sampled by the SAW. For this calculation, it is necessary to remember that the critical percolation fraction for each network is no longer $\frac{1}{2}$ but has a value given by a Gaussian distribution centered around $\frac{1}{2}$. In this case, Levy et al.[37] find reasonable agreement for granular Pb and In films.

Recently, acoustoelectric SAW attention measurements at 168 MHz were performed on an $YBa_2Cu_3O_7$ film deposited on a YZ cut $LiNbO_3$ substrate.[18] The attenuation curve deviates from the resistance curve; there is also a significant difference between the predicted and observed attenuation changes. A scanning electron micrograph of the film shows that the size of the grains is about 3000 Å and that therefore $L \simeq 10$. By using Eqs. (15) and (20) in Ref. 31 and $R_n = (r_1 r_2)^{1/2}$, it is possible to find the minimum resistance of the network r_1 and from this the intrinsic resistivity of the individual grains in the $YBa_2Cu_3O_7$ film. If it is assumed that r_1 measures the resistance of a grain of pure material, the resistivity $\rho = r_1 t_{\mathrm{film}} = 12.5\ \mu\Omega \cdot \mathrm{cm}$, where t_{film} represents the thickness of the film. X-ray analysis of the film has shown that the individual grains are oriented with the c-axis perpendicular to the surface, and therefore both the electric and acoustoelectric resistance measurements are essentially confined to the better conducting ab planes. It is interesting to note that the value obtained for the intrinsic resistance of the grains along the ab planes is at least a factor of 2 lower than what has been measured electrically on the best single crystals of $YBa_2Cu_3O_7$. Therefore, it may be possible that acoustoelectric measurements circumvent any twinning imperfections that may be present in the single crystals which could contribute to the resistivity. Thus, this may be a technique for measuring the average intrinsic resistivity of grains whose dimensions are even less than 3000 Å.

4.3 Quantum Hall Effect

The quantum Hall effect (QHE) gets its name from the fact that the Hall resistance ρ_{xy} of some two-dimensional electron (2DE) systems appears to be quantized in steps to a very high degree of accuracy. The resistance ρ_{xx} of such a system exhibits peaks at the edge of the steps. Since the acoustoelectric attenuation depends on σ_{xx}/σ_{2D}, by evaluating $\sigma_{xx} = \rho_{xx}/(\rho_{xx}^2 + \rho_{xy}^2)$, associating σ_{xx} with $\sigma_\square$, and substituting the result into Eq. (15), one finds that there should be an attenuation minimum at the position of the peak in σ_{xx} with a maximum on either side of the minimum. This is because when $\sigma_{xx}(\text{peak}) > \sigma_{2D}$, the attenuation should go through a maximum on either side of the peaks at the fields at which $\sigma_{xx} = \sigma_{2D}$, resulting in a minimum in attenuation at the position of the peaks in σ_{xx}. This theoretical prediction agrees fairly well with experimental results obtained by Schenstrom et al.[41] using proximity coupling to a III–V semiconducting heterojunction. The observation that the experimental attenuation peaks are broader than the predicted ones provides evidence for delocalization of the quantized electronic states on the scale of the dimensions probed by the 210-MHz SAW, about 2.6 μm.

4.4 Magnetoelastic Interaction

Very large magnetic-field-dependent attenuation of 618 MHz Rayleigh mode SAWs propagating on a quartz substrate through evaporated 200-Å Ni films was discovered in 1976 by Krischer et al.[42] The impressive results described in this work are that a small change in magnetic field from 10 to 90 Oe can produce a change in attenuation of 30 dB/cm and that the maximum in attenuation is located at 10 Oe. The experimentally observed correlation between the attenuation and the in-plane magnetic susceptibility of the Ni films[43] led to a phenomenological model for the observed attenuation response.[44,45] The films are characterized as having in-plane uniaxial anisotropy with single-domain magnetization that changes in direction, through coherent spin rotation, when an external magnetic bias field is applied.

Energy dissipation is produced by magnetoelastic

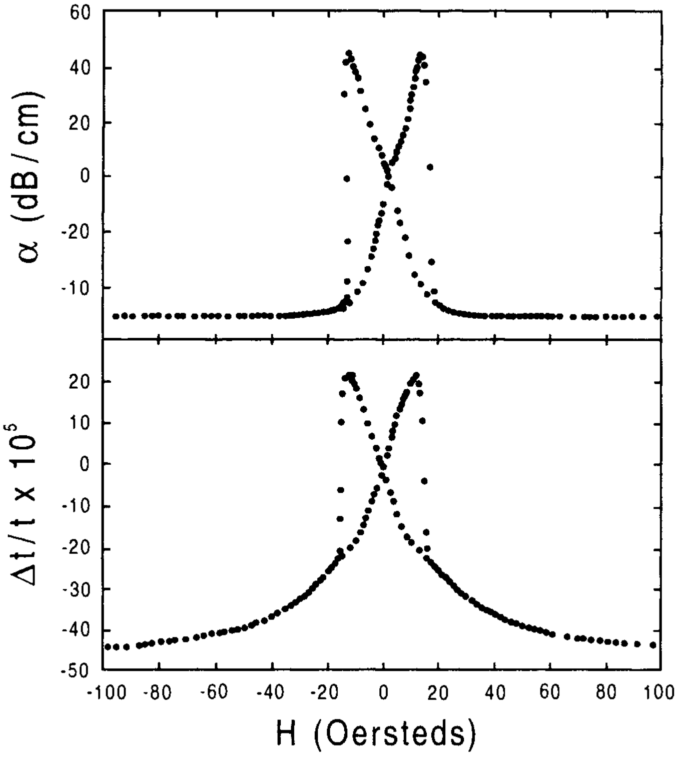


Fig. 5 SAW attenuation and transit time (velocity) as a function of applied magnetic field produced by a 372-Å nickel film deposited on a quartz substrate.[47] The SAW attenuation data are shown at the top and transit time data is displayed at the bottom.

interaction. The time-dependent strains ϵ of the SAW produce oscillating magnetic fields via magnetoelastic interaction. These small magnetic fields oscillate the film magnetization about its equilibrium orientation determined by the larger DC magnetic field. Since there is a finite relaxation time for oscillations of the magnetization, the magnetization does not follow the SAW in phase. This results in energy dissipation per unit volume per cycle given by[45]

$$E_d = \frac{\pi[(\Delta H_A/2)\sin(2\theta_0 - 2\phi)]^2[B^2 + H_s^2(1 + G^2)]H_s MG}{[AB - H_s^2(1 + G^2)]^2 + [H_s G(A + B)]^2}, \tag{17}$$

where H_A is the uniaxial anistropy field that acts along the film's easy axis of magnetization; M is the magnetization; H is the applied field; θ and ϕ are, respectively, the angular positions of M and H_A with respect to H; G is the Gilbert damping parameter; $A = H_A \cos(2\theta_0 - 2\phi) + H \cos\theta_0, B = H\cos\theta_0 + 4\pi M, H_s = \omega/\gamma$, where γ is the electronic magnetomechanical ratio; and $\Delta H_A = \eta_m \epsilon$, where η_m is the magnetoelastic coupling coefficient of the film. The term A represents the restoring force on M in the film plane, while B represents the restoring force on M perpendicular to the film plane. The change in velocity is given by[46] a similar expression.

The shapes of the curves obtained from Eq. (17) and the one for the velocity changes look remarkably close to those displayed in Fig. 5 for the attenuation and transit time (velocity) produced by a 372-Å nickel film deposited on a quartz substrate,[47] where the maximum attenuation change is almost 80 dB/cm. This attenuation change is large enough to be used to develop frequency tunable SAW dispersion lines and filters.[48]

In summary, SAW attenuation measurements provide an invaluable experimental tool with which to probe relevant electrical and physical properties of thin films. Surface acoustic wave attenuation in thin films is an exciting and fruitful area of research whose potential for device application is only beginning to be tapped.

REFERENCES

1. B. A. Auld, *Acoustic Fields and Waves in Solids, Vols. I and II*, 2nd ed., R. E. Krieger, Malabar, FL, 1990.
2. G. S. Kino, *Acoustic Waves: Devices, Imaging and Analog Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1987.
3. D. P. Morgan, *Surface Wave Devices for Signal Processing*, Elsevier, Amsterdam, 1985.
4. S. Datta, *Surface Acoustic Wave Devices*, Prentice-Hall, Englewood Cliffs, NJ, 1986.
5. G. W. Farnell and E. L. Adler, in W. P. Mason and R. N. Thurston (Eds.), *Physical Acoustics*, Vol. IX, Academic, New York, 1972, pp. 35–127.
6. H. E. Bommel, "Ultrasonic Attenuation in Superconducting Lead," *Phys. Rev.*, Vol. 96, 1954, p. 220.
7. W. P. Mason, "Ultrasonic Attenuation Due to Lattice-Electron Interaction in Normal Conducting Metals," *Phys. Rev.*, Vol. 97, 1955, p. 557; W. P. Mason, *Physical Acoustics and the Properties of Solids*, Van Nostrand, Princeton, NJ, 1958.
8. R. W. Morse, "Ultrasonic Attenuation in Metals by Electron Relaxation," *Phys. Rev.*, Vol. 97, 1955, p. 1716.
9. A. B. Pippard, "Ultrasonic Attenuation in Metals," *Phil. Mag., Ser. 7*, Vol. 46, 1955, p. 1104.
10. F. Akao, "Attenuation of Elastic Surface Waves in Thin Film Superconducting Pb and In at 316 MHz," *Phys. Lett.*, Vol. 30A, 1969, p. 409.
11. E. Krätzig, K. Walther, and W. Shilz, "Investigation of Superconducting Phase Transitions in Pb-films With Acoustic Surface Waves," *Phys. Lett.*, Vol. 30A, 1969, p. 411.
12. D. A. Robinson, K. Maki and M. Levy, "Effects of Superconducting Fluctuations on the Ultrasonic Attenuation in a Thin Aluminum Film," *Phys. Rev. Lett.*, Vol. 34, 1974, p. 709.
13. J. G. Park, P. Tai, and S. Frake, "Ultrasonic Surface Wave Attenuation by Induced Superconductivity in PbCu and PbAg 'Sandwiches'," *J. Phys.*, LT15, Vol. 39, No. C-6, Pt. 1, 1978, p. C6/681.
14. H. P. Fredricksen, M. Levy, J. R. Gavaler, and M. Ashkin, "The Ultrasonic Attenuation of 700 MHz Surface Acoustic Waves in Thin Films of Superconducting NbN," *Phys. Rev.*, Vol. B27, 1983, p. 3065.
15. W. E. Barley and B. J. Marshall, "Ultrasonic Attenuation of Surface Acoustic Waves in Superconducting Zinc," *Phys. Rev.*, Vol. B19, 1979, p. 3467.
16. H. L. Salvo, Jr., H. P. Fredricksen, M. Levy, and J. R. Gavaler, "Ultrasonic Attenuation by a Thin Film of High Transition Temperature Nb_3Ge," *Solid State Comm.*, Vol. 33, 1980, p. 781.
17. H. P. Fredricksen, H. L. Salvo, Jr., M. Levy, R. H. Hammond and T. H. Geballe, "Ultrasonic Attenuation of Surface Acoustic Waves in Thin Film of Superconducting Nb_3Sn," *Phys. Lett.*, Vol. 75A, 1980, p. 389.
18. H. P. Baum, B. K. Sarma, M. Levy, J. Gavaler, and A. Hohler, "SAW Measurements on a Nb Film and an $Y_1Ba_2Cu_3O_7$ Film," *IEEE Trans. Mag.*, Vol. 27, 1991, p. 1280.
19. M. Tachiki, H. Salvo, Jr., D. A. Robinson, and M. Levy, "Ultrasonic Mean Free Path in a Granular Aluminum Film," *Solid State Comm.*, Vol. 17, 1975, p. 635.
20. D. R. Snider, H. P. Fredricksen, and S. C. Schneider, "Surface Acoustic Wave Attenuation by a Thin Film," *J. Appl. Phys.*, Vol. 52, 1981, p. 3215.
21. T. Holstein, "Notes on Electroacoustical Phenomena," Westinghouse Research Memo 60-94698-3-MI7, 1956.

22. A. B. Pippard, "Theory of Ultrasonic Attenuation in Metals and Magneto-acoustic Oscillations," *Proc. Roy. Soc.*, Vol. A257, 1960, p. 165.
23. R. S. Sorbello, "Transport Theory of Ultrasonic Attenuation," *Phys. Stat. Sol. (B)*, Vol. 77, 1976, p. 141.
24. J. B. Bardeen, L. N. Cooper, and J. R. Schrieffer, "Theory of Superconductivity," *Phys. Rev.*, Vol. 108, 1957, p. 1175.
25. M. Levy, "Ultrasonic Attenuation in Superconductors for $ql < 1$," *Phys. Rev.*, Vol. 131, 1963, p. 1497.
26. M. Levy, H. Salvo, Jr., D. A. Robinson, K. Maki, and M. Tachiki, "Surface Acoustic Wave Investigation of Superconducting Films," in J. de Klerk (Ed.), *Proceedings 1976 Ultrasonics Symposium*, IEEE 76 CH 1120-5 SU, IEEE, New York, 1976, p. 633.
27. B. R. Tittmann, "Ultrasonic Attenuation in a Type II Superconducting Alloy," *Phys. Rev.*, Vol. B2, 1970, p. 625.
28. R. Kagiwada, M. Levy, I. Rudnick, H. Kagiwada and K. Maki, "Ultrasonic Attenuation in a Pure Type II Superconductor Near H_{c2}," *Phys. Rev. Lett.*, Vol. 18, 1967, p. 74; R. Kagiwada, M. Levy, and I. Rudnick, "Ultrasonic Attenuation in Type II Superconductors," in A. S. Borovik-Romanov and V. A. Tallin (Eds.), *Proceedings of the 10th International Conference on Low Temperature Physics*, Moscow, 1966.
29. H. P. Fredricksen, H. L. Salvo, Jr., M. Levy, R. H. Hammond and T. H. Geballe, "Ultrasonic Attenuation of Surface Acoustic Waves in a Superconducting Thin Film of Nb_3Sn in an Applied Magnetic Field," in J. de Klerk and B. R. McAvoy, (Eds.), *Proceedings 1979 Ultrasonics Symposium*, IEEE 79 CH 1482-9 SU, IEEE, New York, 1979, p. 435.
30. K. A. Ingebrigtsen, "Linear and Nonlinear Attenuation of Acoustic Surface Waves in a Piezoelectric Coated With a Semiconducting Film," *J. Appl. Phys.*, Vol. 41, 1970, pp. 454–459.
31. R. Adler, "Simple Theory of Acoustic Amplification," *IEEE Trans. Sonics Ultrasonics*, Vol. SU-18, 1971, p. 115.
32. A. H. Schenstrom, "Proximity Coupling of Surface Acoustic Waves to Quasi 2-Dimensional Systems," Ph.D. Dissertation, University of Wisconsin-Milwaukee, 1987.
33. J. Schmidt, M. Levy and A. F. Hebard, "Ultrasonic Investigation of Amorphous Superconducting Films," *Phys. Rev.*, Vol. B50, 1994, pp. 3988–3994.
34. L. G. Aslamazov and A. I. Larkin, "Effect of Fluctuations on the Properties of a Superconductor Above the Critical Temperature," *Fiz. Tverd. Tela*, Vol. 10, 1968, p. 1104 (*Sov. Phys. Solid State*, Vol. 10, 1968, p. 875).
35. J. M. Kosterlitz and D. J. Thouless, "Ordering, Metastability and Phase Transitions in Two Dimensional Systems," *J. Phys.*, Vol. C6, 1973, p. 1181.
36. H. Tejima, J. Schmidt, C. Figura, and M. Levy, "Surface Acoustic Wave Investigation of Granular Lead Films," in B. R. McAvoy (Ed.), *IEEE 1983 Ultrasonics Symposium Proceedings*, 83 CH 1947-1, IEEE, New York, 1983, p. 1100.
37. M. Levy, J. Schmidt, A. Schenstrom, M. Revzen, A. Ron, B. Shapiro, and C. G. Kuper, "Surface Acoustic Attenuation in Granular Lead Films," *Phys. Rev.*, Vol. B34, 1986, p. 1508; C. G. Kuper, M. Levy, M. Revzen, A. Ron, and B. Shapiro, "Surface Acoustic Attenuation in Granular Lead Films," in U. Eckern, A. Schmid, W. Weber, and H. Wahl (Eds.), *Proceedings of the 17th International Conference on Low Temperature Physics*, Elsevier Science Publishers, B.V., 1984, p. 913.
38. V. Ambegaokar, B. I. Halperin and J. S. Langer, "Hopping Conductivity in Disordered Systems," *Phys. Rev.*, Vol. B4, 1971, p. 2612.
39. For a discussion of the effective medium approximation, see, S. Kirkpatrick, "Classical Transport in Disordered Media: Scaling and Effective-Medium Theories," *Phys. Rev. Lett.*, Vol. 27, 1971, p. 1722.
40. J. Schmidt, M. Levy, and A. F. Hebard, "Ultrasonic Investigation of Granular Superconducting Films," *Phys. Rev.*, Vol. B43, 1991, p. 505.
41. A. Schenstrom, B. K. Sarma, M. Levy, and H. Morkoc, "Frequency Dependent Breakdown of the Dissipationless State in the Quantum Hall Effect," *Solid State Comm.*, Vol. 68, 1988, p. 357.
42. C. Krischer, I. Feng, J. B. Block, and M. Levy, "Magnetic Field-Dependent Attenuation of Surface Waves by Nickel Thin Films," *Appl. Phys. Lett.*, Vol. 29, 1976, p. 76.
43. I. Feng, H. Fredricksen, C. Krischer, M. Tachiki and M. Levy, "Correlation between the Magnetic Susceptibility of Nickel Films and Their Effect on Surface Acoustic Waves," in J. de Klerk and B. R. McAvoy, (Eds.), *Proceedings 1977 IEEE Ultrasonics Symposium*, 77CH 1264-1 SU, 1977, p. 328.
44. I. Feng, "Interaction of Surface Waves with Thin Nickel Films," Ph.D. Dissertation, University of Wisconsin-Milwaukee, 1979.
45. I. Feng, M. Tachiki, C. Krischer, and M. Levy, "Mechanism for the Interaction of Surface Waves with 200 Å Nickel Films," *J. Appl. Phys.*, Vol. 53, 1982, p. 177.
46. D. R. Walikainen, "Ultrasonic Surface Acoustic Wave Investigation of Thin Magnetostrictive Films," Ph.D. Dissertation, University of Wisconsin-Milwaukee, 1990.
47. R. F. Wiegert and M. Levy, "Enhanced Magnetically Tunable Attenuation and Relative Velocity of 0.6 GHz Rayleigh Waves in Nickel Thin Films," *Appl. Phys. Lett.*, Vol. 54, 1989, p. 995.
48. M. Levy, R. G. Wiegert, B. R. McAvoy, H. L. Salvo, Jr., and D. Bailey, "Variable Bandwidth SAW Filter Using Magnetoelastic Coupling," in B. R. McAvoy (Ed.), *IEEE 1985 Ultrasonics Symposium Proceedings*, 85 CH 2209-S, IEEE, New York, 1985, p. 88.

59

WAVE MODES IN LIQUID HELIUM

GARY A. WILLIAMS

1 INTRODUCTION

The acoustic properties of liquid helium display many unusual characteristics when the liquid is cooled to low temperatures. A phase transition into a superfluid state occurs in which a portion of the fluid is able to flow without friction, having zero viscosity. The flow field associated with this superfluid component introduces new hydrodynamic variables into the equations of motion of the fluid and leads to a rich variety of new sound modes. Studies of these modes have provided a means of measuring a number of the fundamental properties of superfluid helium. There have also been applications of superfluid acoustics to other more general topics in condensed matter physics, such as the characteristics of phonon propagation and critical properties at phase transitions. There are also more practical applications to subjects such as acoustic microscopy and the structure of porous materials.

There are two stable isotopes of helium, ^{3}He and ^{4}He, and both become superfluid at low enough temperatures. The more common isotope ^{4}He is a spin-zero boson, whose superfluid transition temperature is the relatively accessible value $T_\lambda = 2.172$ K at the saturated vapor pressure (SVP). The rare isotope ^{3}He, on the other hand, has a nuclear spin of $\frac{1}{2}$ and forms a Fermi liquid, with the Pauli restriction of two atoms per quantum state preventing the macroscopic occupation necessary for the formation of a superfluid state until very low temperatures are achieved. It is only at temperatures below 1.05 mK at SVP that a BCS type of superfluid transition is observed to occur, where Cooper pairs of the ^{3}He atoms are formed.

In addition to the pure liquids ^{3}He and ^{4}He, superfluidity is also found in liquid mixtures of the two isotopes. In this case it is the ^{4}He that is superfluid, with the ^{3}He remaining a normal fluid and reducing the overall superfluid fraction. There have been searches at ultralow temperatures for a state where both species might simultaneously be superfluid, but at least for temperatures down to 0.1 mK this does not seem to occur.

2 LIQUID ^{4}HE

The phase diagram of liquid ^{4}He is shown in Fig. 1*a*. For temperatures and pressures below the critical point (T = 5.2 K, p = 2.2 bars) helium condenses into a clear, light fluid of density $\rho = 0.145$ g/cm^3 at $T = 0$. By pumping on the vapor above the liquid, the temperature can be lowered, and the normal fluid to superfluid transition at T_λ is marked by a sharp peak in the specific heat that resembles the Greek letter λ, giving rise to the name *lambda transition*. There is also a very large increase in the thermal conductivity at that temperature, causing the noisy boiling of the liquid to abruptly cease. When external pressure is applied to the liquid, the value of T_λ is reduced, reaching a value of $T_\lambda = 1.81$ K at $p = 29$ bars. At pressures above this the liquid then finally solidifies.

2.1 Two-Fluid Model

To describe superfluid ^{4}He, Landau proposed a hydrodynamic formulation[1,2] in which the liquid is taken to be a mixture of two interpenetrating fluids: a superfluid component of density ρ_s and a normal fluid component of density ρ_n. These densities add together to give the measured density of the liquid, $\rho_s + \rho_n = \rho$, and are temperature-dependent, as shown in Fig. 2. The superfluid density ρ_s is zero at T_λ but then increases with fur-

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker
Copyright © 1997 John Wiley & Sons, Inc.

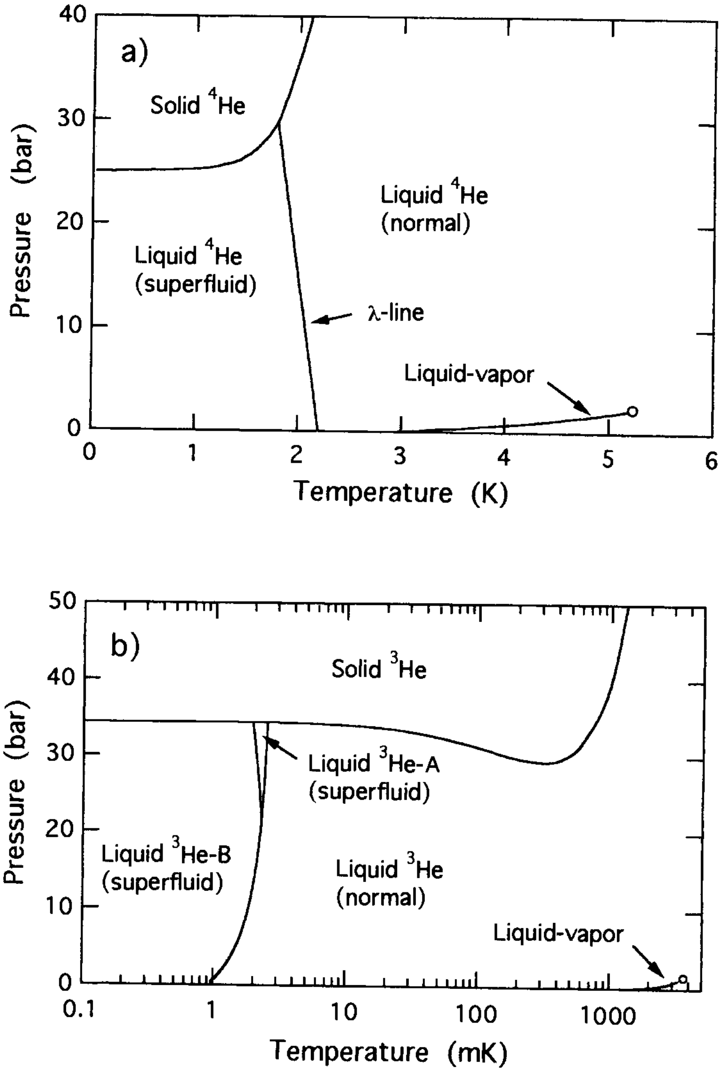


Fig. 1 Phase diagram of (*a*) ^{4}He and (*b*) ^{3}He.

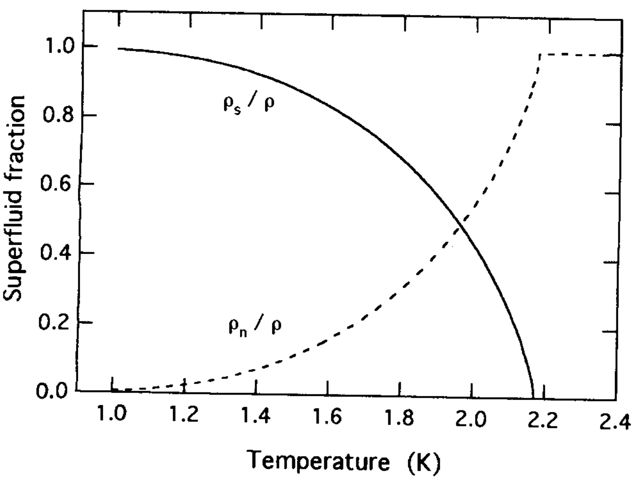


Fig. 2 Superfluid and normal fractions of ^{4}He at SVP.

ther cooling to $\rho_s/\rho = 1$ at $T = 0$, while the behavior of the normal fluid is the converse, with $\rho_n/\rho = 1$ at T_λ, falling to zero at $T = 0$. The viscosity and entropy of the superfluid component are identically zero, a consequence of the superfluidity arising from a macroscopic number of ^{4}He atoms condensing into the lowest quantum mechanical ground state. The normal fluid component, on the other hand, has a finite viscosity η and specific entropy s just like any other fluid. Landau showed that the normal fluid is actually made up of the gas of thermal excitations in the helium, the phonons and rotons.[2]

The superfluid and normal components have their own independent velocity fields $\mathbf{v}_s$ and $\mathbf{v}_n$, and the net mass current is $\mathbf{J} = \rho_s \mathbf{v}_s + \rho_n \mathbf{v}_n$. Landau proposed a set of eight equations of motion for the eight independent variables $\mathbf{v}_s$, $\mathbf{v}_n$, ρ, and s. Conservation of mass yields

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho_s \mathbf{v}_s + \rho_n \mathbf{v}_n) = 0. \tag{1}$$

Since entropy is only carried by the normal fluid, conservation of entropy gives the equation (neglecting dissipative effects such as viscosity and thermal conductivity)

$$\frac{\partial (\rho s)}{\partial t} + \nabla \cdot (\rho s \mathbf{v}_n) = 0. \tag{2}$$

The superfluid component is accelerated by gradients in the chemical potential μ of the liquid,

$$\frac{\partial \mathbf{v}_s}{\partial t} = -\nabla \mu = -\frac{\nabla p}{\rho} + s \nabla T. \tag{3}$$

The fact that the superfluid responds to a temperature gradient is unusual, and it is this property that gives rise to propagating thermal waves. The final equation of motion is that the net mass current in the liquid is accelerated by a pressure gradient:

$$\frac{\partial}{\partial t} (\rho_s \mathbf{v}_s + \rho_n \mathbf{v}_n) = -\nabla p. \tag{4}$$

In the above equations second-order terms in the velocities have been neglected. If we further make the approximations of linear acoustics by taking $\rho = \rho_0 + \delta\rho, p = p_0 + \delta p, s = s_0 + \delta s$, and $T = T_0 + \delta T$, where $\delta\rho \ll \rho_0$, and so forth, then the above equations reduce to a dispersion relation for the phase velocities $c = \omega/k$ of two propagating sound modes,[1]

$$\left(1 - \frac{c^2}{c_1^2}\right)\left(1 - \frac{c^2}{c_2^2}\right) = 1 - \frac{C_p}{C_v}, \tag{5}$$

where $C_p = T(\partial s/\partial T)_p$ and $C_v = T(\partial s/\partial T)_v$, are the specific heats at constant pressure and constant volume, and the velocities c_1 and c_2, discussed in the next two sections, are given by

$$c_1^2 = \left(\frac{\partial p}{\partial \rho}\right)_s \qquad c_2^2 = \frac{\rho_s T s^2}{\rho_n C_p}. \tag{6}$$

2.2 First Sound

Since for liquid ^{4}He the factor $1 - C_p/C_v$ (proportional to the thermal expansion coefficient) is less than 10^{-3}, Eq. (5) to a very good approximation has as its solution two uncoupled sound modes of velocities c_1 and c_2. The first sound mode, with a velocity c_1 determined by the adiabatic bulk modulus, is completely analogous to ordinary sound in a liquid. The normal fluid and superfluid move in phase, with $\mathbf{v}_s = \mathbf{v}_n$, which gives rise to pressure and density oscillations that can be generated and detected with usual acoustic transducers. Figure 3 shows the variation of c_1 with temperature[3] at SVP. The variation in the superfluid regime is a fairly slow decrease from the $T = 0$ velocity of about 237 m/s, except in the region of the λ-point where there is a sharp cusp that is related to the critical behavior of the thermodynamic variables near the transition temperature.[1] With applied pressure the sound velocity increases, reaching a value of 361 m/s at $p = 25$ bars and $T = 0$.

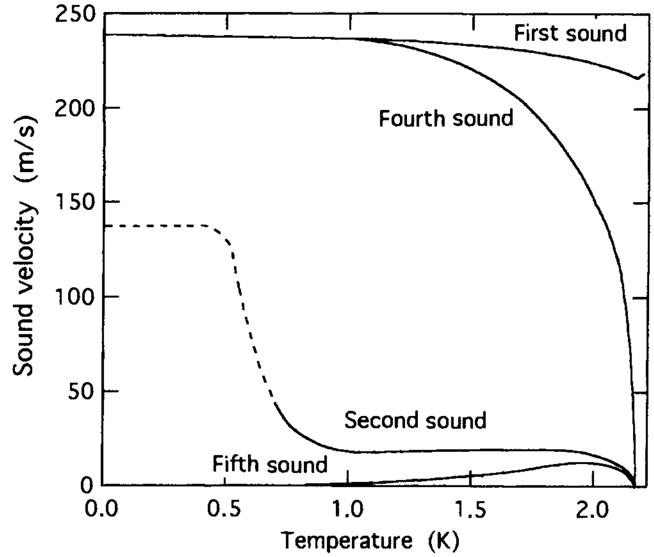


Fig. 3 Velocities of the various sound modes in ^{4}He at SVP.

The attenuation of first sound is shown in Figure 4. Above T_λ the attenuation of first sound is small and is due to the usual viscous and thermal conduction mechanisms found in ordinary fluids. In the superfluid phase the attenuation is dominated by two large peaks, one at T_λ and the other near 1 K. The attenuation peak[5] at T_λ results from the critical-point divergence of the superfluid relaxation time, which varies as $\tau_s = 1.9 \times 10^{-12}|T_\lambda - T|^{-(1.07)}$ (in seconds). The attenuation maximum occurs just slightly below T_λ, at the point where $\omega\tau_s = 1$, with ω the angular sound frequency. The attenuation near 1 K also arises from a relaxation process, in this case the finite relaxation time for the photon and roton excitations of the normal fluid to come into equilibrium with each other. Below 1 K the attenuation falls off rapidly, varying as T.[4] The very low attenuation in this regime allowed the operation of an acoustic microscope in the superfluid[6] at frequencies up to 10 GHz, giving an unprecedented microscope resolution of order 200 Å.

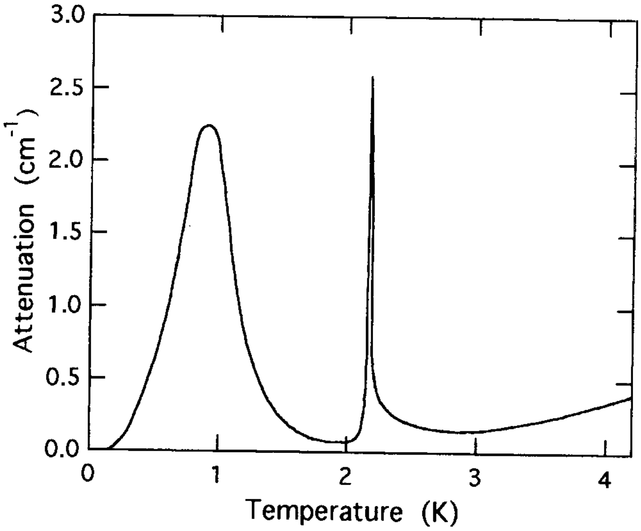


Fig. 4 Attenuation of first sound in ^{4}He at 12 Mhz and SVP. (From Ref. 2.)

2.3 Second Sound

The second solution of Eq. (5), in the limit of a vanishing expansion coefficient, propagates with phase velocity c_2 given by Eq. (6). This mode, second sound, is unique to the superfluid state. It is a propagating temperature and entropy wave, which is unusual because in an ordinary liquid these variables are diffusive. The normal fluid and superfluid move in opposite directions such that the net mass flow $\mathbf{J} = 0$, and hence $\mathbf{v}_s = -(\rho_n/\rho_s)\mathbf{v}_n$. This means that the pressure and density oscillations of the mode are nearly zero, and it is almost entirely a temperature oscillation. In the Landau excitation picture the second sound corresponds to a modulation of the density of the phonon and roton excitations.[2]

To excite and detect second sound, it is necessary to use thermal transducers, such as a heater wire to generate the sound and a resistance bolometer to detect it. Alternatively, it is possible to use an oscillating membrane transducer having small pores in the membrane, such as filter paper. Only the superfluid component with its zero viscosity can flow through the pores, and the oscillating membrane can then produce (or detect) the finite value of $\mathbf{v}_s - \mathbf{v}_n$ which characterizes the second sound mode. The velocity of the second sound can be measured by time-of-flight techniques or by using a resonant cavity. For a cavity the boundary conditions at a solid wall follow from making the assumption that the heat flow from the liquid into a solid wall is zero, $\dot{Q} = 0$, giving the condition $\nabla \delta T = 0$ perpendicular to a wall. For a long closed tube of length L this results in resonant frequencies given by $f_n = nc_2/2L$, where n is an integer.

Figure 3 shows the velocity c_2 at SVP. The low-temperature limit of $c_2 = c_1/\sqrt{3}$ is unobservable in practice (shown by the dashed line) because the excitation mean free path exceeds the second-sound wavelength for temperatures below 0.8 K. In the regime above 1.2 K, c_2 has a maximum value of about 20.4 m/s near 1.65 K, and then falls to zero at T_λ because of the factor of ρ_s in Eq. (6). Under pressure the velocity decreases somewhat,[3] reaching a maximum value at 25 bars of 15.4 m/s at 1.45 K.

Because the thermodynamic quantities s and C_p can be independently determined to good accuracy, measurements of c_2 can be used to extract values of ρ_s. This is particularly useful in the critical regime near T_λ where ρ_s scales as a power law of the temperature difference from T_λ, $\rho_s/\rho = A(T_\lambda - T)^\nu$. Experiments[7] using second sound have found $A = 2.4$ and $\nu = 0.67$, in agreement with theoretical predictions.

The attenuation of second sound is primarily from irreversible heat flow in the temperature oscillations, due to energy transport by the phonon and roton excitations. In pure ^{4}He below 1 K their mean free paths become quite long, and the resulting high attenuation makes the mode unobservable below aboaut 0.8 K. Above 1 K this attenuation mechanism drops rapidly with increasing temperature, until the λ-point is approached. Very close to T_λ the attenuation begins to increase again, and finally diverges as $(T_\lambda - T)^{-0.33}$, due to the diverging superfluid relaxation time at the critical point.

Second sound has also been used to measure the properties of ^{3}He–^{4}He mixtures. At low temperatures in mixtures the normal fluid fraction remains finite since it is proportional to the ^{3}He concentration, $\rho_n/\rho = m_3^* X/m_4$, where m_3^* is the effective mass of a ^{3}He atom and X is the molar concentration of the ^{3}He. The second-sound velocity at low temperatures reduces to $c_2^2 = 5k_BT/3m_3^*$ since the ^{3}He forms a nearly ideal gas of quasiparticles, and measurements[8] in that regime can then determine m_3^*, found to be about 2.3 ^{3}He masses at zero pressure.

2.4 Third Sound

Third sound is a thickness wave that propagates in very thin films of liquid helium.[9] If He gas is admitted to a chamber at low temperatures, an attractive van der Waals potential acts to pull the helium atoms toward the solid surfaces of the chamber. For an atom a distance z from a surface, the potential is $V(z) = -\alpha/z^3$, where α characterizes the dipole–dipole interaction of the atom and the solid surface and has been measured for a number of substrates. The attraction causes a thin liquid film to build up on the substrate to a thickness d where the van der Waals potential equals the chemical potential of the gas,

$$d = \left[\frac{\alpha}{k_BT\ln(p_0/p)}\right]^{1/3}, \qquad (7)$$

where p is the gas pressure in the chamber and p_0 is the saturated vapor pressure of bulk liquid He. This thickness can range from less than a monolayer at the lowest pressures (one layer is a thickness of 3.6 Å) to around 100 layers as p approaches p_0. It has been found that even films as thin as two layers can still be superfluid, and that the two-fluid hydrodynamic description remains valid. Since the films are so thin the normal fluid is viscously clamped to the substrate, $\mathbf{v}_n = 0$, and thus only the superfluid component can participate in the wave motion. The restoring force for oscillations of the thickness of the film is the van der Waals force, and the third sound velocity for a film of thickness d is given by

$$c_3^2 = \frac{\rho_s}{\rho} \frac{3\alpha(d - D_0)}{d^4}. \tag{8}$$

Because of the solidification of the first layer or so near the substrate and also depletion of the superfluid density from the proximity of the substrate, not all of the thickness of the film is superfluid. This is accounted for in Eq. (8) by the effective nonsuperfluid thickness D_0, which is subtracted from d. Figure 5 shows the third-sound velocity as a function of film thickness at three different temperatures. The substrate used for this data[10] was a plate of calcium fluoride. The thickness oscillations of third sound are accompanied by temperature oscillations (mounds are colder since they have more superfluid flowing into them, and depressions are hotter), and so heater wires can serve as generators of third sound. Receiver transducers can either be resistance bolometers or capacitance detectors of the film height. At the thinnest films in Fig. 5 the third-sound signal abruptly disappears, and this is due to the sharp rise in attenuation at the onset of the Kosterlitz–Thouless superfluid phase transition in the two-dimensional film.[11]

2.5 Fourth Sound

Fourth sound is a pressure-density wave that propagates in helium contained in a porous superleak material.[12] The pores of the superleak must be small enough that the normal fluid is viscously locked to the substrate and only the superfluid participates in the wave motion. This requires that the pores be smaller than the viscous penetration depth $\ell_v = \sqrt{2\eta/\rho_n\omega}$, where η is the normal fluid viscosity and ω the angular frequency of the wave. Typically for frequencies in the kilohertz range this means pores smaller than several micrometres.

The velocity of fourth sound in an ideal superleak, such as straight capillary channels, is given by[1]

$$c_4^2 = \frac{\rho_s}{\rho} c_1^2 + \frac{\rho_n}{\rho} c_2^2. \tag{9}$$

The second term on the right of Eq. (9) is very small compared to the first and can generally be neglected. The connection to second sound, however, does show that

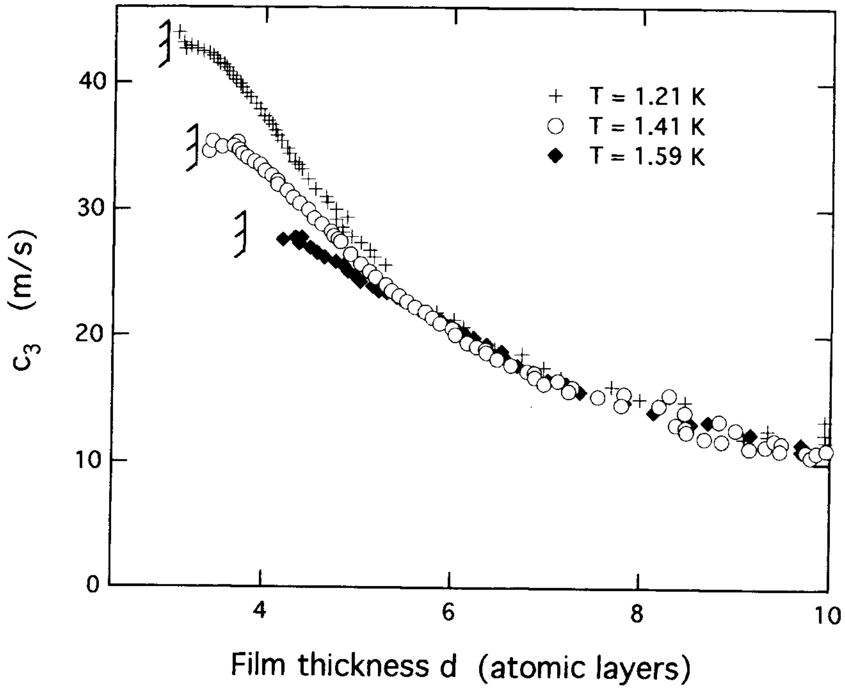


Fig. 5 Velocity of third sound on a CaF$_2$ substrate. (From Ref. 10, but with their thickness scale divided by 1.48 to account for a revised van der Waals constant.)

fourth sound also has an oscillating temperature component and can be detected with either pressure transducers or thermal transducers. Figure 3 shows the fourth-sound velocity, which drops to zero at T_λ because of the factor ρ_s/ρ. In fact the fourth-sound velocity is one of the most accurate techniques for measuring the superfluid fraction,[3] since both c_1 and c_4 can be measured to high precision.

The attenuation of fourth sound is typically very small when the pore size is much smaller than the viscous length. As the pore size is increased, the attenuation increases, reaches a maximum, and then decreases when the pores become larger than the viscous length. In this regime the mode velocity changes from that of fourth sound to first sound as the normal fluid unlocks and begins to participate in the wave motion.

In an actual superleak material with randomly arranged grains, the fourth-sound speed is lower than that given by Eq. (9), with the experimentally measured speed being characterized by an index of refraction n, $c_{4\text{exp}} = c_4/n$. The speed is lowered because of the tortuous path that the fluid has to follow through the pores. The index is related to the tortuosity parameter α of the Biot theory for hydrodynamics in porous media[13] by $n^2 = \alpha$. The index depends on the porosity ϕ of the porous material, which is the ratio of the open-pore volume to the total volume of the sample. Experiments[13] find that the index for porosities not too small follows the variation predicted by Archie's law, $n^2 = b\phi^{-m}$, where b is a constant (≈ 1) and m is an exponent that depends on the grain shape; $m = \frac{1}{2}$ is predicted for spherical grains. At very low porosities there are deviations from Archie's law: Experiments in sintered porous ceramics[14] show a diverging tortuosity near the percolation threshold (at $\phi \approx 0.06$ in these samples) where there is no longer a connected acoustic path through the ceramic.

2.6 Fifth Sound

Fifth sound is a temperature wave that propagates in a superleak where a pressure-release boundary condition prevents fourth sound from propagating.[15] It is analogous to second sound, except that the normal fluid is viscously clamped by the superleak, $\mathrm{v}_n = 0$. The velocity of the fifth sound is given by

$$c_5^2 = \frac{\rho_n}{\rho} c_2^2. \tag{10}$$

This is the pressure-release limit $c_1 \to 0$ of Eq. (9). Experimentally, the pressure-release condition is obtained by having a free surface present, that is, only partially filling the superleak with liquid. This eliminates the bulk modulus contribution to the sound speed, although there are still small surface tension and van der Waals restoring forces that must be subtracted to isolate the fifth-sound term.

Figure 3 shows the predicted c_5 as a function of temperature, and Fig. 6 shows the comparison with experimental measurements. In pure ^{4}He c_5 is zero at $T = 0$ where $\rho_n/\rho = 0$. With increasing temperature it increases to a maximum of 12 m/s at 1.8 K, and then falls to zero at T_λ since $c_2 \to 0$. In a ^{3}He–^{4}He mixture c_5 remains finite as $T \to 0$ since ρ_n/ρ from the ^{3}He remains finite at $T = 0$. The experimental speed of fifth sound in a porous material is also subject to a lowered velocity from an index of refraction, with the additional feature that the

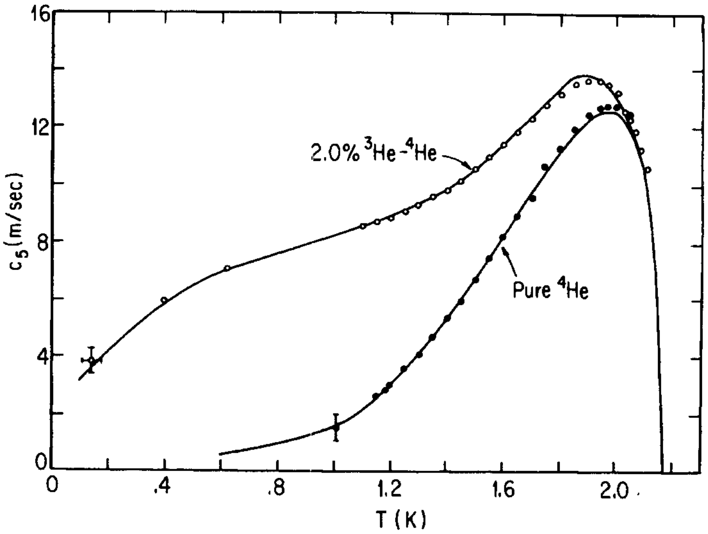


Fig. 6 Velocity of fifth sound in ^{4}He and ^{3}He-^{4}He mixtures. (From Ref. 15.)

index is found to be a function of the amount of liquid in the partially filled superleak.[15]

2.7 Two-Phase Sound

Two-phase sound is a mode that propagates in an acoustic waveguide partially filled with superfluid.[16] The sound in the vapor above the liquid helium couples to the second sound in the superfluid, resulting in a single propagating mode characterized by both temperature and pressure oscillations. The phase velocity of the waveguide mode, shown in Fig. 7, depends on the depth L of the liquid in the waveguide, and varies smoothly from the vapor sound velocity c_v when $L = 0$ to the second-sound velocity c_2 when the waveguide is completely filled to $L = H$, where H is the height of the waveguide. The waveguide geometry is shown in the inset of Fig. 7, where the propagation direction is out of the page. The theoretical prediction for the phase velocity of the mode is

$$c^2 = \frac{c_2^2\left(1+\dfrac{c_6^2}{c_2^2}\,\dfrac{H-L}{L}\right)}{\left(1+\dfrac{c_6^2}{c_v^2}\,\dfrac{H-L}{L}\right)}, \qquad (11)$$

where

$$c_6^2 = \frac{\left(s_v\,\dfrac{\partial p}{\partial T}\right)_{\mathrm{co}}}{\left(\rho\,\dfrac{\partial s}{\partial T}\right)_p} \qquad (12)$$

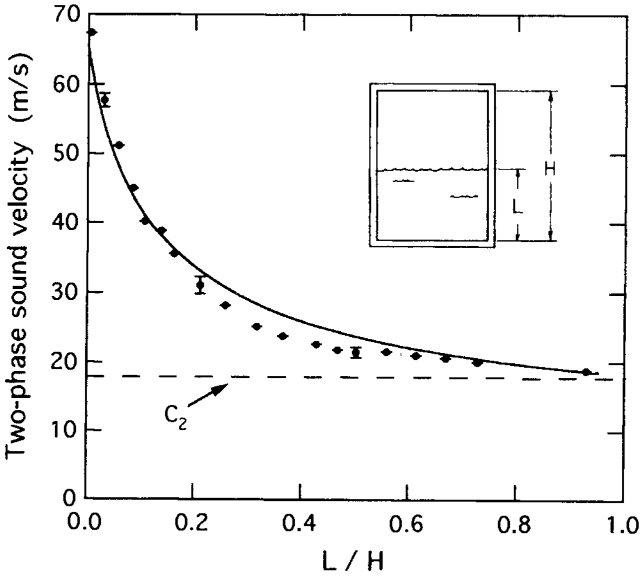


Fig. 7 Velocity of two-phase sound in ^{4}He as a function of the liquid height in the waveguide. (From Ref. 16.)

is a velocity characteristic of the liquid–vapor coupling and where s_v is the specific entropy of the vapor and $(\partial p/\partial T)_{\mathrm{co}}$ is the temperature derivative of the liquid–vapor coexistence curve. A second-sound temperature swing δT in the liquid gives rise to a pressure swing δp given by the coexistence curve, $\delta p = (\partial p/\partial T)_{\mathrm{co}}\delta T$; this pressure increase in the vapor gives an additional restoring force to the mode that results in a phase velocity greater than c_2, as seen in Fig. 7.

At finite frequencies the attenuation of two-phase sound is quite strong, arising from the nonequilibrium nature of the evaporation–condensation process at the free surface. A theory of the attenuation using kinetic models for the nonequilibrium Onsager coefficients agrees well with experimental measurements.[16]

3 LIQUID ^{3}HE

The isotope ^{3}He can be liquified at 3.2 K at 1 bar of pressure. Less dense than ^{4}He ($\rho = 0.08$ g/cm^3 at $T = 0$, $p = 0$), ^{3}He is a liquid of particles obeying Fermi–Dirac statistics (since the atoms have a nuclear spin of $\frac{1}{2}$) with an effective Fermi temperature $T_F = 1.6$ K at zero pressure. At temperatures well below T_F the liquid displays the characteristics of a weakly interacting degenerate Fermi gas, having a heat capacity linear in T, a temperature-independent spin susceptibility, and a viscosity that varies as $1/T^2$. It is not an ideal gas, however, and the ideal-gas expressions for thermodynamic quantities are modified due to the interactions between the quasiparticles at the Fermi surface. Laudau formulated a theory for this Fermi liquid that has been successful in providing a phenomenological description,[17] with the various interaction parameters of the theory being determined from experimental measurements.

Figure 1*b* shows the phase diagram of liquid ^{3}He. The Fermi statistics does not allow macroscopic occupation of the ground state as occurs in ^{4}He, and this greatly suppresses the occurrence of superfluidity, to temperatures only below a few millikelvins. The transition that occurs there is a BCS pairing transition, in which attractive interactions (primarily arising from spin–spin fluctuations) between quasiparticles near the Fermi surface cause them to form Cooper pairs and create a superfluid condensate. Unlike superconductors where the electron pairs have angular momentum $L = 0$ and spin $S = 0$, the ^{3}He pairs form with $L = 1$ and $S = 1$. There are two different superfluid phases,[17] the A and B phases shown in Fig. 1, which differ in the components of the triplet spin wave function of the pairs. In the B phase all three components of the pair spin wave function are present: $\uparrow\uparrow$, $\downarrow\downarrow$, and $\uparrow\downarrow$. In the A phase the amplitude of the $\uparrow\downarrow$ component is zero. This gives rise to a strong anisotropy in many of the properties of the A phase, determined with respect to a unit vector **l** along the direction of the angular momentum of the Cooper pairs. The direction of **l** is affected by magnetic fields, by boundary walls, and by flow of the liquid. One of the important anisotropic quantities is the superfluid density, which becomes a tensor. If measured along the direction of **l**, it has a value $\rho_{s\parallel}$, whereas in a direction perpendicular to **l** it has a different value $\rho_{s\perp}$ that is larger by about a factor of 2. In the B phase the superfluid density is isotropic, making its superfluid properties more similar to the case of ^{4}He.

3.1 First and Zero Sound

The velocity of first sound in normal ^{3}He in the low-temperature regime where it is a degenerate Fermi liquid is predicted by the Landau theory to be

$$c_1^2 = \tfrac{1}{3}v_F^2(1 + F_0), \tag{13}$$

where

$$v_F = \frac{1}{m_3^*}\left(3\pi^2\hbar^3 \frac{N}{V}\right)^{1/3}$$

is the Fermi velocity, with N/V the particle density and m_3^* the quasiparticle effective mass; F_0 is one of the interaction parameters of the Landau theory, and since m_3^* can be obtained from specific heat measurements, F_0 can be determined from experimental measurements[17-19] of c_1. At 0 bar c_1 = 185 m/s, yielding $m_3^*/m_3 = 2.8$ and $F_0 = 9.3$, and increases with pressure to c_1 = 320 m/s at 34 bars, where $m_3^*/m_3 = 5.8$ and $F_0 = 87$.

The above expression for c_1 is valid only in the limit of low frequencies, where the period of the sound wave is much larger than the quasiparticle scattering time τ, so that $\omega\tau \ll 1$. Since the occupation of quasiparticles at the Fermi surface and the empty hole states for them to scatter into are proportional to T, the scattering time varies as $\tau \approx 1 \times 10^{-6}/T^2$ (s/mK^2). At megahertz frequencies and sufficiently low temperature the sound characteristics undergo a change to a collisionless regime where $\omega\tau \gg 1$. In this regime the mode is called zero sound, with a velocity that is a few percent higher than first sound.[18] Figure 8 shows the change in sound velocity with temperature as the mode changes from first sound to zero sound. At the crossover point $\omega\tau \approx 1$ there is a peak in the attenuation, also shown in Fig. 8. The change in velocity and attenuation can be modeled in terms of a theory of viscoelasticity[20] with a single relaxation time τ, which for the velocity and attenuation gives

$$c = c_1 + (c_0 - c_1)\left(\frac{\omega^2\tau}{1 + \omega^2\tau^2}\right), \tag{14}$$

$$\alpha = \frac{c_0 - c_1}{c_1^2}\left(\frac{\omega^2\tau}{1 + \omega^2\tau^2}\right), \tag{15}$$

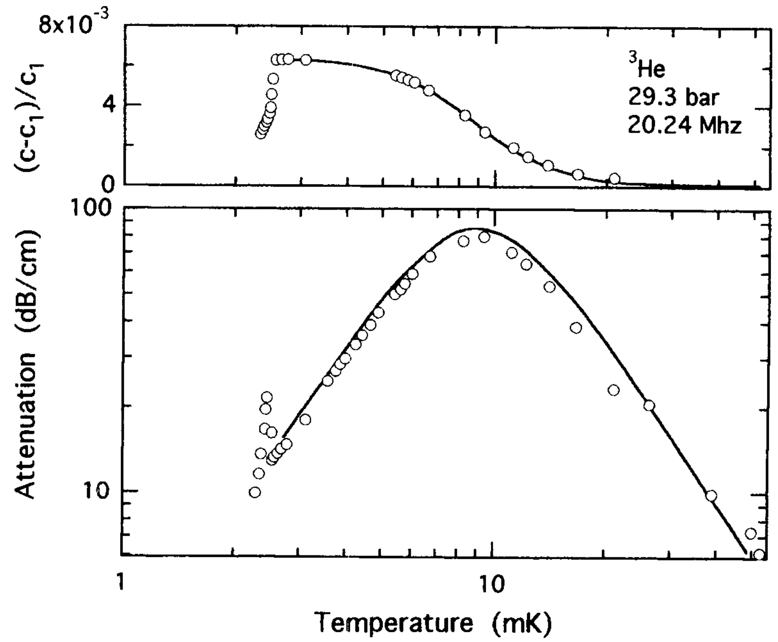


Fig. 8 Velocity and attenuation of sound in liquid ^{3}He, showing the first sound to zero sound crossover and the superfluid onset near 2.5 mK at this pressure. (From Ref. 18.)

where c_1 is the first-sound velocity and $c_0 = \sqrt{c_1^2 + (4K_\infty/3\rho)}$ is the zero-sound velocity with K_∞ the viscoelastic shear modulus at infinity frequency. In the collisionless regime the shear viscosity becomes a shear restoring force, as reflected in the increased velocity and decreased attenuation for zero sound. The shear restoring force also means that there will be a propagating shear wave with velocity $c_s = \sqrt{4K_\infty/3\rho}$, which is known as transverse zero sound. This has been observed directly using quartz shear wave transducers,[21] with the overdamped viscous shear waves in the high-temperature regime turning into the propagating transverse zero sound in the low-temperature collisionless regime.

In the superfluid part of the phase diagram the zero sound displays a number of unusual features. As seen in Fig. 8, at temperatures just below T_c the zero-sound velocity drops from c_0 back to c_1, and a number of peaks in the sound attenuation are observed as the ^{3}He is cooled to low temperatures.[18, 19, 22] In the B phase these peaks correspond to mode crossings of the zero sound with excited states of the Cooper pairs, when the energy $\hbar\omega$ of the sound wave matches the energy of the excited state. Because of the complexity of the ^{3}He order parameter, there are many possible states, but most of them do not couple to density oscillations. Two that do couple strongly have been dubbed the "squashing mode" (a pair state with $J = 2$, $J_z = 2$, and an imaginary amplitude) and the "real squashing mode" (with $J = 2$, $J_z = 2$, and a real amplitude). These are observed at sound energies $\hbar\omega_{\text{sq}} = 0.82\Delta(T)$ and $\hbar\omega_{\text{rsq}} = 0.59\Delta(T)$, in reasonable agreement with theoretical predictions, where $\Delta(T)$ is one-half of the pair energy gap. In the region of the mode crossings a very strong dispersion of the phase velocity of the sound is observed, as expected for sound propagation in a resonant medium, and there is also a strong distortion of the shape of a received sound pulse.[19]

In the A phase there are also attenuation peaks from the excited modes of the order parameter, and the anistropy of the order parameter means that this attenuation will be strongly dependent on the angle between the sound propagation **q** and the vector **l** describing the Cooper pair orientation. The direction of **l** can be varied with a magnetic field, and measurements[22] of the anistropy of the attenuation agree well with theoretical predictions.[18]

3.2 Second Sound

Second sound has been observed in superfluid ^{3}He, but the experiments are generally quite difficult because of very high attenuation. In the isotropic B phase the second-sound velocity is given by Eq. (6). Because the temperature T in this equation is only a millikelvin or so for superfluid ^{3}He and yet the specific heat C_p is still large, the velocity is quite low compared to the case of ^{4}He, being only a few centimeters/second. The combination of a low sound velocity and high viscosity η leads to a very high damping of the mode, since the attenuation is proportional to η/c_2^3. Despite this, a very broad low-frequency resonance corresponding to second sound was observed[23] in ^{3}He-B.

Second sound has been more definitively observed[24] in the A_1 phase of ^{3}He. This is a phase that appears in a narrow temperature range about T_c when a strong magnetic field is applied to the liquid. A field of $0.1T$ gives a temperature width of about 6 μK for the A_1 phase. In this phase the magnetic Zeeman energy makes it favorable for only the $\uparrow\uparrow$ Cooper pairs to form. With only one spin type paired, an oscillation in the superfluid density (Cooper pair density) is then also an oscillation in the spin density. The restoring force for a wave of such oscillations then includes a magnetic force that is much larger than the thermal forces of usual second sound, and the wave is really a combination of second sound and a spin density wave. The velocity is given by

$$c_2^2 = \frac{\rho_s}{\rho_n}\frac{\rho}{\chi}\left(\frac{\gamma\hbar^2}{2m_3}\right), \tag{16}$$

where χ is the magnetic susceptibility of the liquid and γ is the gyromagnetic factor. A typical value of the velocity is of order 1 m/s, and the measurements[24] agreed well with the above theoretical expression. Because of the higher velocity, the attenuation was much smaller than for second sound in the B phase, giving a well-defined propagating mode.

3.3 Fourth Sound

The observation of a propagating fourth-sound mode in liquid ^{3}He was one of the first proofs that the liquid in fact had undergone a superfluid transition.[25] By using Eq. (9) the superfluid density was obtained from the measured resonant frequencies, and it was found to obey a linear dependence on the reduced temperature, $\rho_s \propto 1 - T/T_c$, as predicted by theory.[17] The initial measurements were performed in packed powder of rather large grain size, about 50 μm. More recent measurements[25] have been carried out in aluminum powders of smaller size, 1–3 μm. In these powders the pore size is comparable to the coherence length of the Cooper pairs, which ranges from about 500 Å at 0 bar to 200 Å at 20 bars. At the pore wall the superfluid density falls to zero, and then away from the wall it increases toward the bulk value over a coherence length. In the small pores there will thus be a considerable reduction of the superfluid density from its bulk value, and this was observed in the fourth-sound measurements.[26] By comparing the reduced superfluid density with the known superfluid density of bulk ^{3}He-B, values of the coherence length could be extracted from the measurements, in reasonable agreement with theory.

REFERENCES

1. S. J. Putterman, *Superfluid Hydrodynamics*, North-Holland, Amsterdam, 1974.
2. K. R. Atkins, *Liquid Helium*, Cambridge Univ. Press, Cambridge, 1959; J. Wilks, *The Properties of Liquid and Solid Helium*, Oxford Univ. Press, Oxford, 1967.
3. J. Heiserman, J. P. Hulin, J. Maynard, and I. Rudnick, "Precision Sound-Velocity Measurements in He II," *Phys. Rev. B*, Vol. 14, 1976, p. 3862; J. Maynard, "Determination of the Thermodynamics of He II from Sound-Velocity Data," *Phys. Rev. B*, Vol. 14, 1976, p. 3868.
4. M. Barmatz and I. Rudnick, "Velocity and Attenuation of First Sound near the λ Point of Helium," *Phys. Rev.*, Vol. 170, 1968, p. 224.
5. R. Carey, C. Buchal, and F. Pobell, "Attenuation and Dispersion of First Sound near the Superfluid Transition of Pressurized ^{4}He," *Phys. Rev. B*, Vol. 16, 1977, p. 3133.
6. J. Heiserman and C. Quate, "The Scanning Acoustic Microscope," in D. Sette, (Ed.), *Proc. of Int. School Enrico Fermi*, Course XCIII, North-Holland, Amsterdam, 1986, p. 343.
7. L. Goldner, N. Mulders, and G. Ahlers, "Second Sound Very Near T_λ," *J. Low Temp. Phys.*, Vol. 93, 1993, p. 131.
8. D. S. Greywall and M. A. Paalanen, "Measurements of Second Sound in Partially Spin-Polarized ^{3}He–^{4}He Solutions," *Phys. Rev. Lett.*, Vol. 46, 1981, p. 1292.
9. K. R. Atkins and I. Rudnick, "Third Sound," in C. J. Gorter (Ed.), *Progress in Low Temperature Physics*, North-Holland, Amsterdam, 1970, Vol. 6, p. 37.
10. I. Rudnick, R. S. Kagiwada, J. C. Fraser, and E. Guyon, "Third Sound in Adsorbed Superfluid Films," *Phys. Rev. Lett.*, Vol. 20, 1968, p. 430.
11. I. Rudnick, "Critical Surface Density of the Superfluid Component in ^{4}He Films," *Phys. Rev. Lett.*, Vol. 40, 1978, p. 1454; V. Ambegaokar, B. Halperin, D. Nelson, and E. Siggia, "Dynamics of Superfluid Films," *Phys. Rev. B*, Vol. 21, 1980, p. 1806.
12. K. A. Shapiro and I. Rudnick, "Experimental Determination of the Fourth Sound Velocity in Helium II," *Phys. Rev.*, Vol. 137, 1965, p. A1383.
13. D. L. Johnson, T. J. Plona, C. Scala, F. Pasierb, and H. Kojima, "Tortuosity and Acoustic Slow Waves," *Phys. Rev. Lett.*, Vol. 49, 1982, p. 1840.
14. M. Bernard and G. A. Williams, "Surface and Bulk Tortuosity of Porous Ceramics near the Percolation Threshold," *Phys. Rev. Lett.*, Vol. 67, 1992, p. 2585.
15. G. A. Williams, R. Rosenbaum, and I. Rudnick, "Fifth Sound in Superfluid ^{4}He and ^{3}He–^{4}He Mixtures," *Phys. Rev. Lett.*, Vol. 42, 1979, p. 1282; G. J. Jelatis, J. A. Roth, and J. D. Maynard, "Observation of Fifth Sound in a Planar Superfluid ^{4}He Film," *Phys. Rev. Lett.*, Vol. 42, 1979, p. 1285.
16. S. Putterman, D. Heckerman, R. Rosenbaum, and G. A. Williams, "Superfluid Two-Phase Sound," *Phys. Rev. Lett.*, Vol. 42, 1979, p. 580; G. A. Williams, "Attenuation of Superfluid Two-Phase Sound," *J. Low Temp. Phys.*, Vol. 50, 1983, p. 455.
17. A. J. Leggett, "A Theoretical Description of the New Phases of Liquid ^{3}He," *Rev. Mod. Phys.*, Vol. 47, 1975, p. 331; J. Wheatley, "Experimental Properties of Superfluid ^{3}He," *Rev. Mod. Phys.*, Vol. 47, 1975, p. 415; *Helium Three*, W. P. Halperin and L. P. Pitaevski (Eds.), North-Holland, Amsterdam, 1990.
18. P. Wolfle, "Sound Propagation and Kinetic Coefficients in Superfluid ^{3}He," in D. Brewer (Ed.), *Progress in Low Temperature Physics*, North-Holland, Amsterdam, 1978, Vol. 7A, p. 191.
19. W. P. Halperin and E. Varoquaux, "Order-Parameter Collective Modes in Superfluid ^{3}He," in W. P. Halperin and L. P. Pitaevski (Eds.), *Helium Three*, North-Holland, Amsterdam, 1990, chap. 7, p. 353.
20. I. Rudnick, "Zero Sound and the Viscoelasticity of Liquid ^{3}He," *J. Low. Temp. Phys.*, Vol. 40, 1980, p. 287.
21. P. R. Roach and J. B. Ketterson, "Observation of Transverse Zero Sound in Normal ^{3}He," *Phys. Rev. Lett.*, Vol. 36, 1976, p. 736.
22. P. R. Roach, B. M. Abraham, P. D. Roach, and J. B. Ketterson, "Anisotropy of the Propagation of Sound in the A Phase of Superfluid ^{3}He," *Phys. Rev. Lett.*, Vol. 34, 1975, p. 715.
23. S. T. Lu and H. Kojima, "Observation of Second Sound in Superfluid ^{3}He–B," *Phys. Rev. Lett.*, Vol. 55, 1985. p. 1677.
24. L. R. Corrucini and D. D. Osheroff, "Observation of Second Sound in ^{3}He–A$_1$," *Phys. Rev. Lett.*, Vol. 45, 1980, p. 2029.
25. H. Kojima, D. N. Paulson, and J. C. Wheatley, "Propagation of Fourth Sound in Superfluid ^{3}He," *Phys. Rev. Lett.*, Vol. 32, 1974, p. 141.
26. K. Ichikawa, S. Yamasaki, H. Akimoto, T. Kodama, T. Shigi, and H. Kojima, "The Superfluid Fraction of ^{3}He Confined in Pores of Sintered Silver," *Phys. Rev. Lett.*, Vol. 58, 1987, p. 1949; T. Hall, S. Tholen, K. Lane, V. Kotsubo, and J. Parpia, "Healing Length of Superfluid ^{3}He," *J. Low Temp Phys.*, Vol. 89, 1992, p. 897.

60

ULTRASONIC INSTRUMENTS FOR NONDESTRUCTIVE TESTING

EMMANUEL P. PAPADAKIS

1 INTRODUCTION

This chapter deals with ultrasound as used in nondestructive testing (NDT). Ultrasonics as limited to NDT in this chapter can be defined as actively generated low-intensity mechanical radiation above human hearing that can be directed for interrogation purposes. As ultrasonic waves propagate in materials, they are influenced not only by the flaws[1] but also by the mechanical properties[2–5] and can be used to determine some of them nondestructively.[6] Since NDT is such a results-oriented and ad hoc interdisciplinary field, it is appropriate to focus on ultrasonic instruments. To conduct actual nondestructive testing one does not interlace a pulse generator, a piezoelectric plate, an amplifier, and an oscilloscope with BNC cables as in a laboratory; one buys a portable, battery-operated flaw detector with built-in pulser, amplifier, oscilloscope, touch-pad controls, and computerized functions that will fit down the conning tower of a submarine.[7] The only cable goes to the potted transducer. For a different kind of job, one may install a large system with manipulators and multiaxis computer control to move sensors (transducers) around in an immersion tank as large as two lanes of an Olympic swimming pool.[8] To get an idea of size and proportion, see Fig. 1. Succinctly, the result desired determines the system configuration, while the system is self-contained and specialized.

The self-contained and specialized instruments all operate to make measurements on the basis of a few fundamental principles. The fundamentals will be treated first in a generic way such that the reader can refer to other chapters in this handbook for details. Then certain special cases will be handled. Special-purpose and general-purpose instruments will be illustrated. A section will be dedicated to a buyers' guide for instruments and accessories by category and function.

2 FUNDAMENTALS

2.1 General

Ultrasound propagates in materials.[9] The propagation is characterized by the mode of the wave,[10] by the variables velocity and attenuation,[11] by the phenomenon of refraction,[12] and by the reflectivity of a discontinuity the wavefront may meet.[13] To make a measurement, the ultrasound must be transmitted and received. The latter is the process of transduction in which a device called a transducer changes an electrical variable (voltage or current in some waveform) into a mechanical wave, and also does the reverse, outputting an electrical signal characteristic of the interaction of the mechanical radiation with the workpiece.[14] To perform an ultrasonic nondestructive measurement on a workpiece, it is necessary to assemble a proper combination of instrument, transducer, coupling (since the waves travel through materials), and possibly focusing, refraction, and mode conversion mechanisms to introduce a propagating wave of a desired mode into the workpiece of a particular geometry.

2.2 Modes

In the medium, the mechanical radiation will propagate in allowed modes.[10] In a bulk solid, the waves can be longitudinal (compressional) and transverse (shear).[2] If the bulk solid is isotropic (the same in all directions),

---

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker
Copyright © 1997 John Wiley & Sons, Inc.

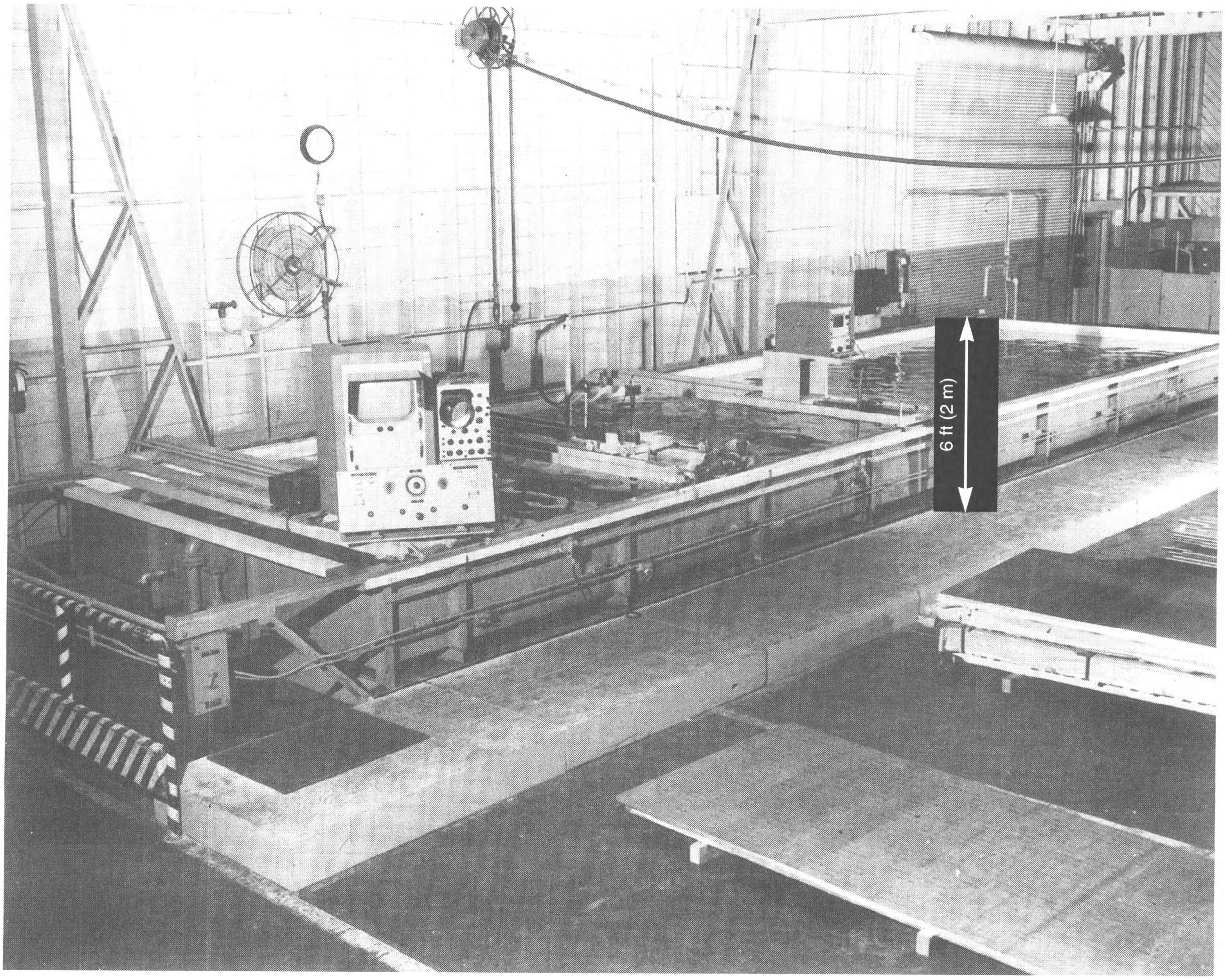


Fig. 1 Immersion tank with ultrasonic scanning equipment showing size and scale of commercial instrumentation. (Courtesy of Douglas Aircraft Company.)

there will be two velocities, one longitudinal and one transverse. If the bulk solid is anisotropic, having directional properties, then there are three velocities (one longitudinal and two shear) with directional variability. (Actually, they are quasi-longitudinal and quasi-shear.) The symmetries of materials can be treated as in crystallography with point groups and elastic moduli.[15] The longitudinal and shear modes (or quasi-longitudinal and quasi-transverse except along pure mode axes) can be used to characterize the workpiece in different ways.

On the surface of a solid, Rayleigh (surface) waves can propagate. Stonley waves run along interfaces.[16]

For plates, the modes are termed Lamb waves (longitudinal and flexural) and Love waves (horizontally polarized shear waves).[17] The Lamb waves have components of particle motion both normal to the plate and in the direction of propagation (in the plane of the plate), while the horizontally polarized shear waves have particle motion in the plate but normal to the direction of propagation. These plate waves are of different orders depending on the number of nodal planes there are in the plate.[18] The most useful wave mode for NDT is the zeroth-order Lamb mode, which can be visualized as a symmetric set of bulges and contractions propagating along the plate (the ostrich swallowing the orange). The useful frequency range for this mode is near the low-frequency limit where the wavelength is much longer than the plate thickness so that the mode is not dispersive. Modes with even numbers of nodal planes (including zero) are flexural.

Waves in wires are similar. There are extensional, flexural, and torsional modes.[19]

2.3 Velocity

The velocity of a wave is determined by the elastic moduli of the medium and the mode of a propagation.[10] As the mode depends on the boundary conditions, the moduli are combined differently to express the velocities of various modes.[20] Fundamentally, however, the dependence of the velocity of a mode upon the elastic moduli of the medium permits one to monitor some of the mechanical properties of the medium.[21] Processing steps that produce desired mechanical properties also change the moduli.[22] Velocity can be used to monitor the process,[21] perform quality assurance,[6] carry out statistical process control,[23] and sort with 100% testing (NDT).[24] If the velocity is known, then travel time measurements yield a thickness gauge with one-sided access.

2.4 Attenuation

The attenuation of a wave is determined by scattering[25] and absorption,[3] which are properties of the medium, and by beam spreading from a finite source.[26] As the scattering and absorption are frequently determined by processing steps, attenuation can be used in quality and process control in some cases.[27] In practice, the measurement of attenuation poses some problems, making its use somewhat limited.

2.5 Reflectivity

All wave modes are partially reflected by changes in the acoustic impedance along their path.[28] Except in some special applications, the reflectivity of air versus a solid can be considered 100%. Reflectivity makes ultrasound ideal for the detection and measurement of cracks and voids in solids.

2.6 Transduction

Transduction from electrical to mechanical energy and from mechanical to electrical energy[14] is performed by specialized, ruggedized devices known as transducers.[29] The most common form is based on an encapsulated longitudinal mode piezoelectric plate.[30] Other types are magnetostrictive, inductive (electromagnetic acoustic transducers), or capacitative. Special front members are used for focusing and refraction with emphasis on mode conversion to get shear waves into solids by means of longitudinal transducers.

2.7 Focusing, Refraction, and Mode Conversion

All three processes depend on refraction. Ultrasound is refracted by Snell's law just as in optics except that a solid medium can support two refracted waves from a single incident wave.[31] These are the longitudinal and shear waves with particle motion in the sagittal plane. A typical use would be to introduce both waves into a solid from a longitudinal wave in a liquid. Often the incident angle is specified so that the refracted angle of the longitudinal wave is beyond the critical angle in the solid, leaving only the vertically polarized shear wave in the solid. Refraction, of course, can be used with lenses for focusing. In going from a liquid into a solid, a wave is bent away from the surface normal. A plano-concave plastic lens accomplishes focusing from a solid transducer into water.[32]

3 ULTRASOUND, ELECTRONICS, AND READ-OUT

3.1 Generic Instrument

What might be termed a "generic ultrasonic instrument" is shown in Fig. 2. Under command of a synchronization generator (at, say, 1000 repetitions per second), the pulser sends electrical energy to the transducer through a pulse limiter circuit, which clips the voltage seen by the amplifier down to a threshold but lets the full voltage hit the transducer. The transducer transmits the ultrasonic wave and receives the echoes. The resulting voltage, not clipped by the pulse limiter because the echo voltage is below the clipping threshold, goes to the amplifier and then to the display. The computer may or may not be present and may simply represent some analog function.

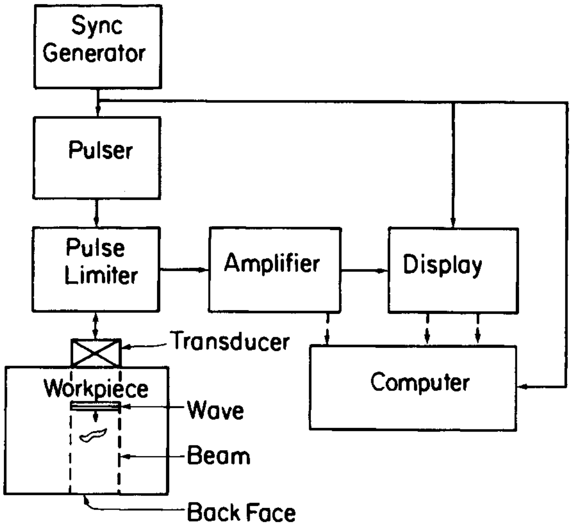


Fig. 2 Generic ultrasonic instrument in block diagram form. A synchronization generator establishes a repetition rate for a pulse input and a display module (with or without computer). The transducer sends ultrasound pulses into the workplace and receives echoes. The electronics allows the echoes to be displayed and manipulated.

3.2 The Generic Display: A-Scan

For the flaw detection instrument with no "intelligence," the display is an oscilloscope with limited adjustability. With well-damped transducers, the resulting display of amplitude versus time (and hence amplitude versus distance in the workpiece) is shown in Fig. 3*a*. With less damped or with electrically tuned transducers, the echoes contain more radio frequency (RF) cycles; usually the result is rectified and detected before display, as in Fig. 3*b*. These sketches could be termed the "generic ultrasonic display." From a technical point of view, they are termed an A-scan, which is simply amplitude versus time for the echoes within the beam of the transducer.

In the A-scan, one sees multiple echoes between the front and back faces of the workpiece plus the echoes from flaws. Beyond the first back face reflection, the echo pattern may become very garbled but may, on the other hand, yield especially interesting data in some cases.[7]

3.3 Analog Modifications to A-Scans for Amplitude

Commercial flaw detection instruments provide various analog processes to help analyze the data. The first, mentioned already, is to rectify and detect the echoes. A second is a baseline suppressor, which eliminates a few decibels of amplitude just above zero amplitude, thus making the more significant flaw echoes easier to discern. A third is a range gate, which picks out a time range of interest from t_1 to t_2 along the A-scan baseline. The signals within the range gate may be either (a) suppressed or (b) analyzed to the exclusion of other signals. The analysis may be by internal computer, external computer, internal alarm analog circuit, or external analog instrument. Note the C-scan discussion below. An alarm can be activated if echoes within the range gate exceed a certain amplitude or threshold.

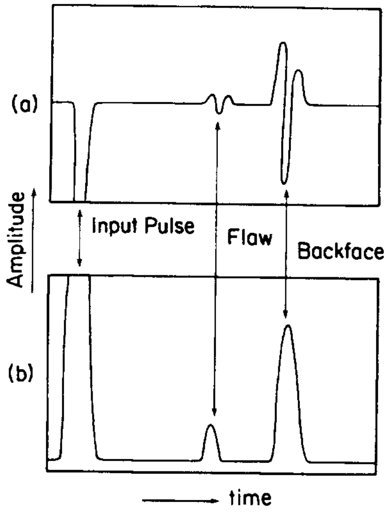


Fig. 3 Generic ultrasonic display for the instrument in Fig. 2. The amplitude is shown, either as RF in (*a*) or as rectified and detected in (*b*), versus time, which is equivalently distance into the workpiece. This display is known as an A-scan.

3.4 Analog Modifications to A-Scans for Time Measurement

The time from the electrical pulsing of the transducer to the arrival of the first back face echo represents either the thickness of the workpiece if the velocity is known or the velocity in the workpiece if the thickness is known. The ultrasonic thickness gauges, which do not require access to the back side of a part, work on the principle of knowing (or at least calibrating for) the velocity. The synchronizing timer starts the charging of a capacitor through a resistor in an *RC* network. The voltage on the capacitor increases as $V = V_0(1 - e^{-t/RC})$, which is close to linear for times such that $V \ll V_0$. The *RC* time constant is built such that this condition holds over the range of time to be measured in one round trip travel time of the wave in the thickest specimen for the given range of calibration, for example, 0.05–1.00 cm, then 1.00–20.00 cm, and so forth. The arrival of the first back-wall reflection triggers the cessation of the charging of the capacitor. The voltage then represents the travel time and hence can be read out on a light-emitting diode (LED) or liquid-crystal display (LCD) as thickness, given the velocity as a multiplier, as noted earlier. If the thickness is known, then the velocity can be measured by varying the velocity-setting dials until the read and known thicknesses match. Obviously, a flaw could vitiate both types of time-dependent measurements. Most thickness gauges are small and portable with no video displays except numerals.

3.5 Mechanical Scanning

General Since a picture is worth a thousand words, developments have been made to provide ultrasonic pictures of the interiors of workpieces. Some sort of mechanism to pick up signals from a number of areas to represent a slice through a part is required. Then a two-dimensional electronic or electromechanical display can provide a picture of the slice.

The most straightforward and simplistic methodology is to scan the transducer over the part or to move the part in front of the transducer with mechanical drives in some sort of scanning pattern. See Fig. 4 where the transducer is shown able to move in the *x*- and *y*-directions while propagation is in the *z*-direction (travel time *t* is along

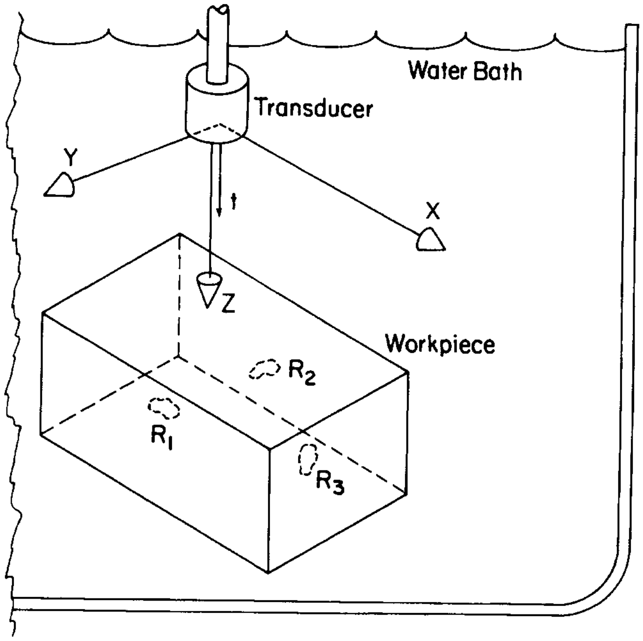


Fig. 4 Schematic diagram of a scanning mechanism in an immersion tank. The transducer can be translated and angulated to aim at a part appropriately. Some display and recording means is always also attached.

the z-direction). There is to be a one-to-one relationship between a point on a display and a region in the workpiece from which an echo comes. How is this achieved? There are two standard scans:

B-Scans A B-scan will portray an x–z slice of the workpiece at a set position Y_i, then make another x–z slice at Y_{i+1}, and so forth. The transducer is moved along x at a rate permitting the ultrasonic pulse to interrogate the workpiece many times over the x-dimension of the workpiece. The pulse repetition rate must be low enough to permit the previous pulse to be attenuated below the noise level before the next pulse is transmitted. The coordinates of the display are the position x_R where the transducer was at the time of arrival t_R of a reflection from a reflector R (R_1, R_2, or R_3 in the figure, depending on whether the beam in slice y passes over them). Then t_R is the other coordinate of the point (x_R, t_R). If the display means has a gray scale, then the brightness of the display corresponds to the amplitude of the reflection. For an on–off scale, a threshold is usually set so that only echoes above a certain amplitude are displayed. The B-scan is used extensively in medical ultrasonics, particularly obstetrics.[33]

C-Scans In the C-scan configuration, the transducer is scanned along x for consecutive steps y_i. However, the coordinate on the display is (x, y) and the display is nonzero only if a reflector R is within the beam along z centered at (x, y). In this configuration, the echoes from the front and back faces would swamp the display, so they are eliminated by setting a range gate to accept and transmit echoes to the display if and only if they arise from volumes below the top face and above the bottom face of the workpiece. Obviously, the range gate can be set narrower such that only signals from a slice thinner in the z-dimension are allowed. Also, a threshold can be set within the gate so that only signals above a certain amplitude will be displayed.[34] Alternately, a gray scale can be added.

For the B-scan as well as the C-scan, a computer can be appended for data storage and image enhancement. Pseudocolor gray scales on color monitors are useful.

3.6 Two Transducers and Other Items

Worth mentioning for completeness are methods where the receiver is a separate transducer. Two configurations are common: pitch and catch and through transmission.

Pitch-and-Catch Configuration Here the two transducers are on the same face of the workpiece, often close together.

Through-Transmission Configuration Here the two transducers are on opposite faces of the workpiece. Often they are in a water tank with the workpiece between them.

Squirters When a tank would be inconvenient, a squirter can be used to convey the ultrasound into and out of the workpiece. A squirter is a laminar flow water hose nozzle with a transducer mounted coaxially inside. Its flow can be aimed at the workpiece to allow normal incidence or refraction at angles.

4 EXAMPLES OF INSTRUMENTS

4.1 General

Following the strategy in Section 1, which delineated flaw detection and materials properties measurement as the two big areas of ultrasonic NDT, both types of instruments will be illustrated. Because flaw detection is the heart of NDT, flaw detection instruments will be treated first.

4.2 Flaw Detection

Flaw detection by ultrasound could be traced back to Langevin's work on piezoelectric crystals for sonar dur-

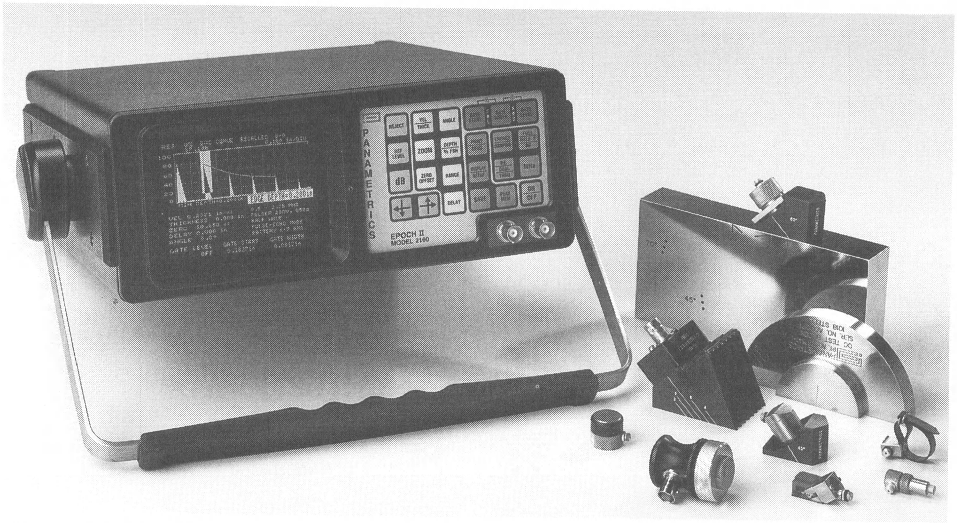


Fig. 5 Modern flaw detection instrument with CRT display. Along side are transducers and calibration blocks. (Courtesy of Panametrics, Inc.)

ing World War I. However, more modern pulse-echo radar when combined with piezoelectric crystals led to pulse–echo ultrasonics in the early 1940s. The first model supersonic reflectoscope by Firestone[35] was as large as a refrigerator. (One will note that the date is early enough to be prior to the division of nomenclature leaving "supersonic" to aerodynamics and "ultrasonic" to mechanical vibrations). A few of these were made and applied in war production. Interestingly enough, just such a flaw detection instrument with a shear wave transducer was instrumental in the first demonstration by Firestone and Fredericks of ultrasonic double refraction in worked metals,[36] a *mechanical property* having to do with differences in shear wave velocity as a function of polarization direction.

Figure 5 shows a modern flaw detection instrument with a cathode ray tube (CRT) as its display. Its pulser is a charged capacitor switched onto the transducer by a charge-conducting diode (CCD); the charge then drains away through a damping resistor. The frequency is broadband and is determined by the highly damped mechanical resonance[37] of the piezoelectric element in the transducer.[30]

The direction of future instruments may be defined by the instrument shown in Fig. 6. In this subcompact piece, the bulk and the weight have been drastically reduced by eliminating the CRT with its high-voltage transformer and substituting a digital LCD display. In the terminology of Fig. 2, all functions beyond the amplifier have been digitized and computerized. All the functions mentioned in Sections 3.3 and 3.4, in addition to rectification mentioned in Section 3.2, have been put into software to

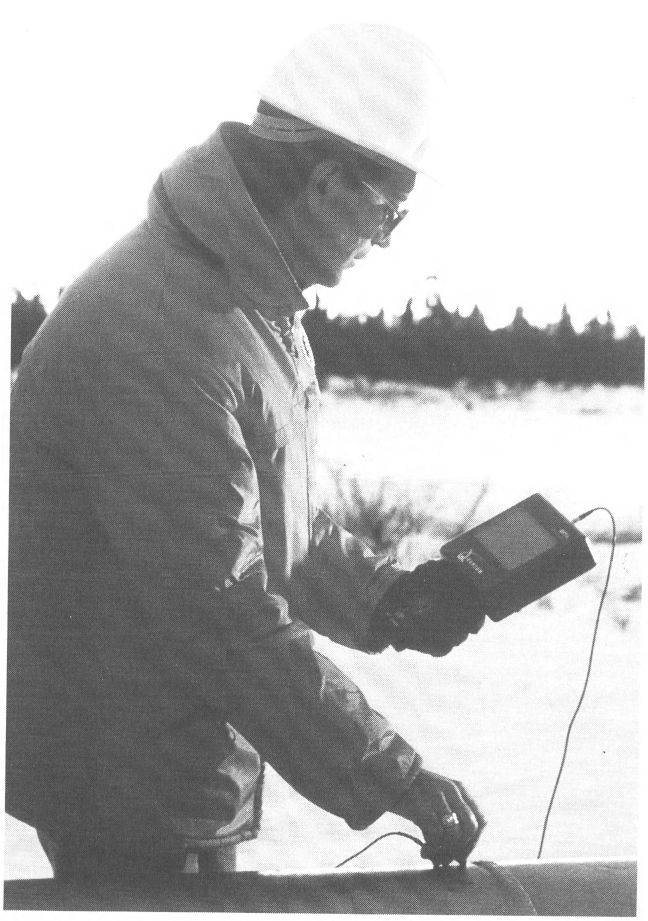


Fig. 6 Subcompact flaw detection instrument with maximum possible digital processing and control and flat LCD display. (Courtesy of DuPont NDT Systems.)

miniaturize the system down to 3 lb (less than 1.5 kg). Controls are also computer operated. The analog section is relegated to the function of acquiring data and "talking" to the computer through an analog-to-digital converter, which takes a digital "snapshot" of what would be the display in Fig. 3a.

Deviating from the tendency to miniaturization is the type of installation that is dedicated to testing huge parts such as aircraft panels and structures. The flaw detection instruments are state-of-the-art, of course. One example of a system utilizing water jets (squirters) on large contoured honeycomb panels is shown in Fig. 7. Scanning motion in this photograph is across the width of the gantry. The jet holders may be rotated 90 degrees to allow scanning along the length of the gantry. Both upper and lower jet holders move in *X*, *Y*, *Z*, and two angular axes to maintain the jets normal to the part surface. The motion of the lower head is controlled by data downloaded from a computer-assisted design (CAD) system. The upper head motion is the mirror image of the lower head, while maintaining a constant spacing between the ends of the jets. Parts within a 20 × 20 × 6 ft envelope (6 × 6 × 2 m) may be scanned with this system.[38]

Much more advanced signal processing methodologies can be called upon to interpret the echoes in terms of the size and shape of the reflector or scatterer. Among these methods are the synthetic aperture focusing technique[39] (SAFT), familiar from radio astronomy, and computed tomography[40] (CT), familiar from x-ray radiography.

4.3 Materials Properties

As soon as ultrasonic waves could be propagated and received in beams by instruments, it was realized that physical properties associated with elasticity could be measured by ultrasound. Leading these experimentalists were T. A. Litovitz,[3] W. P. Mason,[2] and R. Truell.[41] They tackled both ultrasonic attenuation and ultrasonic velocity, breaking new ground that led their disciples to develop measurement methods and instrumental embodiments to determine mechanical properties in practical applications. Their pioneering work led to other engineering devices and to elegant science, as well.

One instrument[42] for measuring ultrasonic velocity to high absolute accuracy[43] is shown in Fig. 8. It employs the pulse–echo overlap method for ultrasonic travel time measurements.[44,45] Under ideal conditions and using calibration of its time base against frequency standards, this instrument can make measurements to an absolute accuracy[43] (not just precision) of ± 5 ppm. For quality assurance of certain hi-tech materials, such as recoined

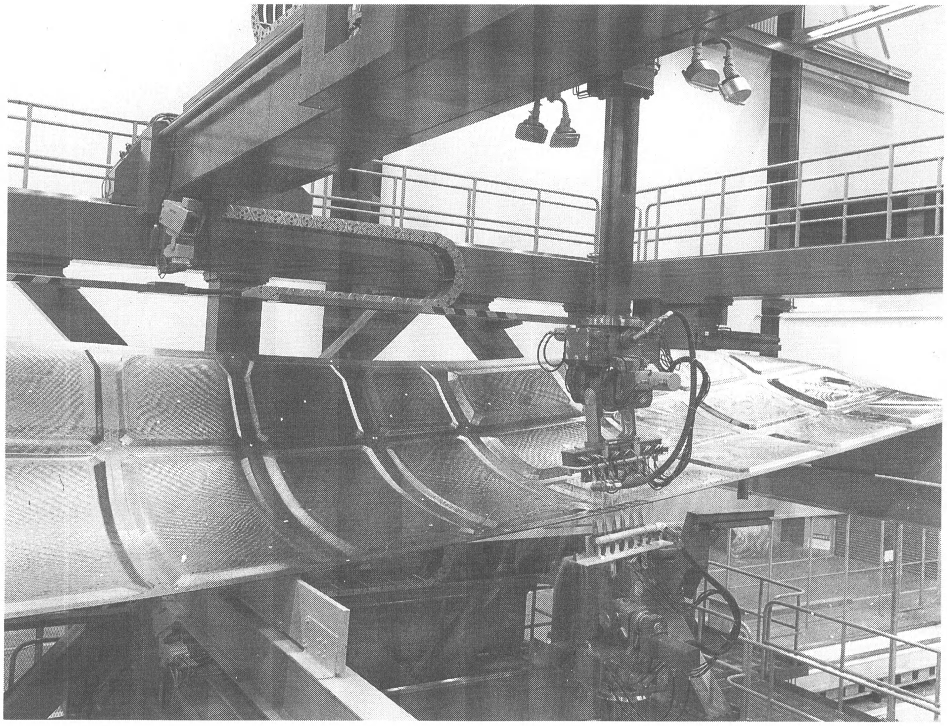


Fig. 7 Large steered scanning ultrasonic test system using squirters to inspect large contoured parts. (Courtesy Boeing Commercial Airplane Division.)

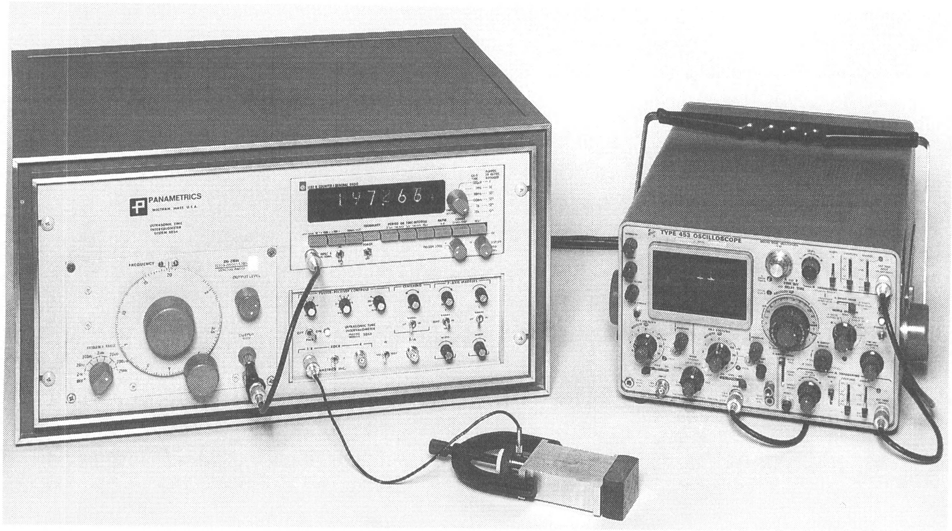


Fig. 8 Commercial pulse–echo overlap instrument for highly accurate ultrasonic velocity measurements. This instrument is used in quality control, scientific experimentation, and development of NDT procedures dependent on ultrasonic velocity. (Courtesy of Panametrics, Inc.)

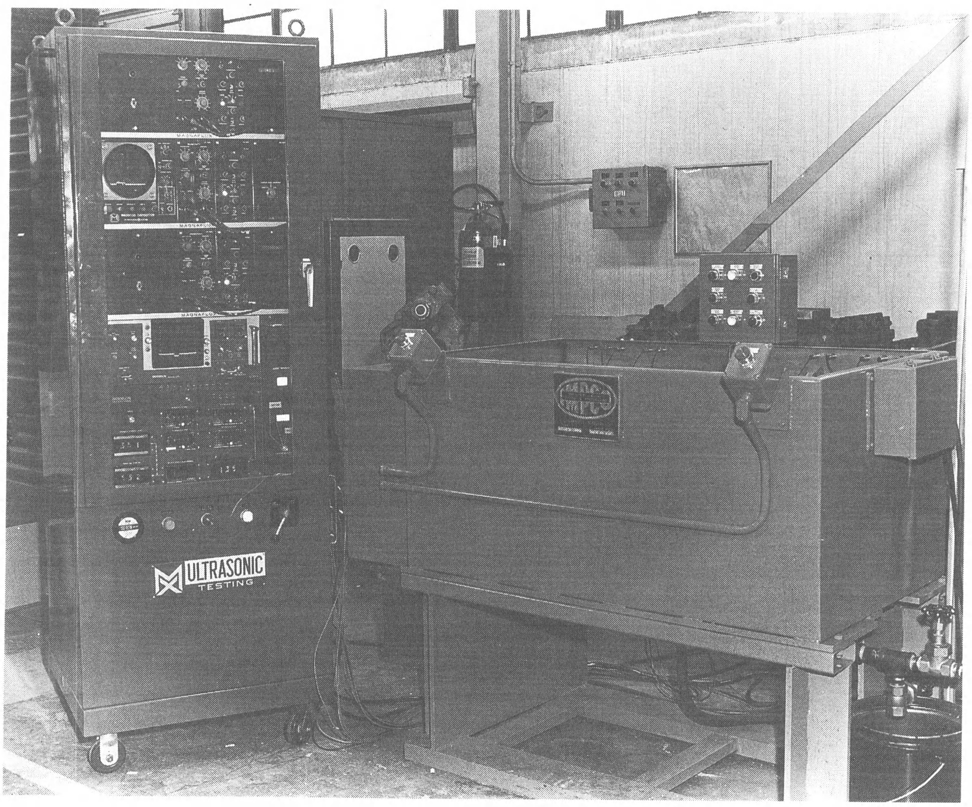


Fig. 9 Ultrasonic system for measuring ultrasonic velocity in modular iron in a foundry for quality assurance. Parts are tested immersed in the tank. (From Ref. 21. Courtesy of Academic Press.)

sintered powder metal parts, high-density ceramics, and vitreous silica, this degree of accuracy is necessary. The laboratory prototype of this instrument was used in quality assurance efforts on vitreous silica for dispersive delay lines[46] in which a few degrees of phase error across the broad ultrasonic beam would cause degradation of performance.[47] The instrument in Fig. 8 is sold worldwide.

Instruments[21] like the ones shown in Figs. 9 and 10 are in actual factory operations throughout the world. This automatic system measures ultrasonic velocity in cast iron to assure quality of nodular cast iron.[21] Molten iron is treated with an additive to cause its carbon to nucleate and grow into microscopic spheres (nodules) when nodular cast iron is the desired product. Nodular iron is a substitute for forged steel in strength and is lower in cost, being cast near-net-shape. Having the 14% by volume of carbon in graphite spheres instead of in graphite flakes or other spread-out forms of graphite maximizes the connectedness of the iron and minimizes stress risers. Thus, the nodular iron has both the maximum strength and the maximum modulus of the possible resultant irons.[48] Maximum modulus implies maximum velocity, so ultrasonic velocity is a quantitative NDT test for the strength of nodular cast iron. Having this quality assurance tool available is necessary because failure modes and effects analysis[49] (FMEA) indicates that improper treatment in the molten state causes deviations from the nodular condition. Improprieties can be caused by "fading" over time before solidification in a batch

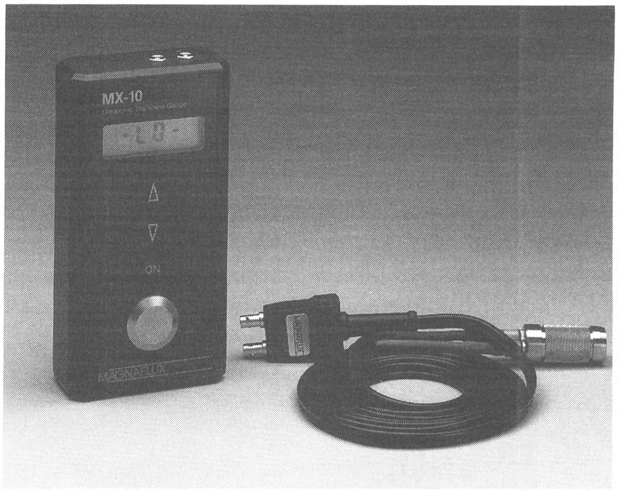


Fig. 11 An ultrasonic thickness gauge about the size of a five-pack of cigars. (Courtesy of Magnaflux, Inc.)

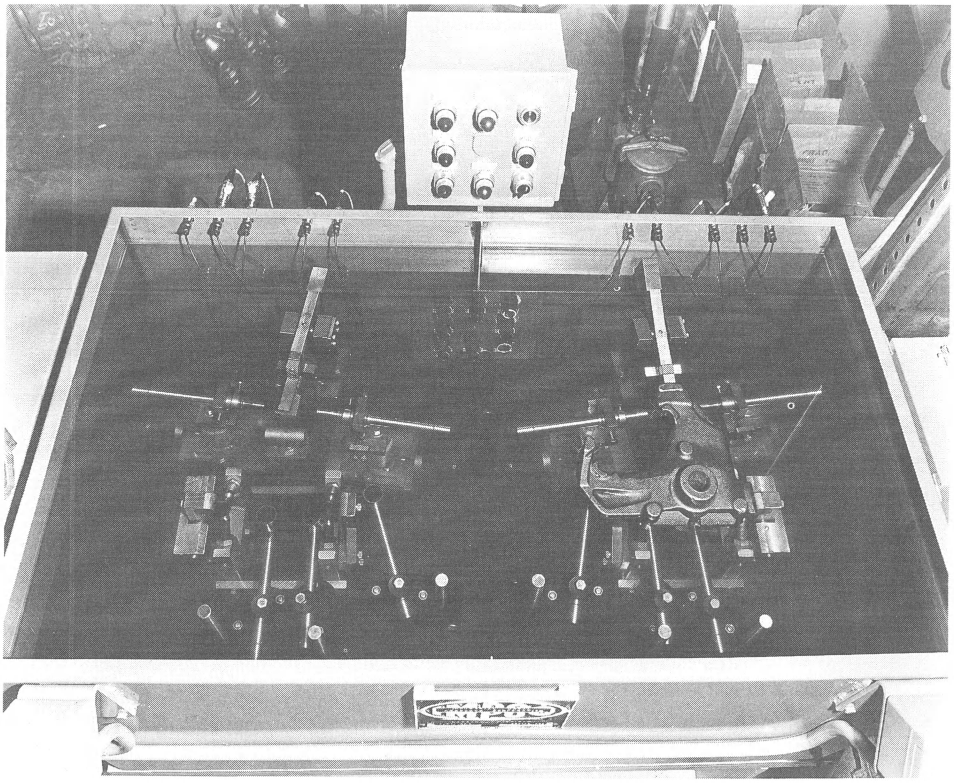


Fig. 10 View of parts in the immersion tank in Fig. 9. One can see the transducers aiming through the part. (From Ref. 21. Courtesy of Academic Press.)

treatment process or else too fast or too slow dissolving of the treatment agent (or too little of it) in an in-mold treatment process. Thus, ultrasound is paying its way in safe car and train parts, to name a few items.

The last item, a thickness gauge, is shown in Fig. 11. This palm-of-the-hand model is one among many that has evolved for practical use where thickness is to be measured from access to one side of the material only. Thickness gauges are placed here under materials properties since the ultrasonic velocity in the workpiece must be known so that the measured travel time can be translated into thickness.

5 BUYER'S GUIDE

The coordinating society for NDT in the United States is the American Society for Nondestructive Testing (ASNT). On a worldwide basis it is one of the top few and probably is the leader. Annually, the ASNT publishes a "Buyer's Guide" in one issue of *Materials Evaluation* (its journal) with a practical operational emphasis. As companies are founded, acquired, divested, and terminated at a rate too great to enter them into a handbook of enduring worth, this section is devoted to listing the categories under which buyers can find ultrasonic testing items in the ASNT "Buyer's Guide".[50] This list may change in future years with editorial policy and technological growth; one can get an idea of the subject from this listing:

Airborne ultrasound
Blocks, test
Bond testers
Bubblers
Calculators, flaw
Couplant, dry
Couplant, wet
Detectors, flaw
Extensometers (bolt gauges)
Flaw discriminators
Gages, thickness
Instruments, multichannel
Manipulators
Processors, signal
Pulsers/receivers
Scanners, lab
Squirters
Systems, computerized
Systems, custom
Systems, EMAT
Systems, handling
Systems, imaging
Systems, piping tubing
Systems, velocity/integrity
Tank, immersion
Testers, concrete
Testers, hardness
Testers, wood
Transducer analyzers
Transducers
Other ultrasonic products

REFERENCES

1. F. A. Firestone, "The Supersonic Reflectoscope for Interior Inspection," *Metals Progress*, Vol. 505, September, 1945.
2. W. P. Mason, *Physical Acoustics and the Properties of Solids*, Van Nostrand, Princeton, 1958.
3. K. F. Herzfeld and T. A. Litovitz, *Absorption and Dispersion of Ultrasonic Waves*, Academic, New York, 1959.
4. D. Setti (Ed.), *"Dispersion and Absorption of Sound by Molecular Processes,"* Course XXVII, in *Proceedings of the International School of Physics "Enrico Fermi,"* Academic, New York, 1963.
5. R. T. Beyer and S. V. Letcher, *Physical Ultrasonics*, Academic, New York, 1969.
6. E. P. Papadakis, "Challenges and Opportunities for Nondestructive Inspection Technology in the High-Volume Durable Goods Industry," *Mat. Evaluation*, Vol. 39, No. 2, 1980, pp. 122–130.
7. T. M. Mansour, "Ultrasonic Inspection of Spot Welds in Thin-Gage Steel," *Mat. Evaluation*, Vol. 46, No. 5, 1988, pp. 640–658.
8. D. J. Hagemaier, "Ultrasonic Maintenance Testing of Aircraft Structures," in *ASNT Handbook*, Vol. 7, American Society for Nondestructive Testing, Columbus, OH, 1991.
9. R. B. Lindsay, *Mechanical Radiation*, McGraw-Hill, New York, 1960.
10. B. A. Auld, *Acoustic Fields and Waves in Solids*, Vols. I and II, Wiley-Interscience, New York, 1973.
11. E. P. Papadakis, "The Measurement of Ultrasonic Velocity" and "The Measurement of Ultrasonic Attenuation" in R. N. Thurston and A. D. Pierce (Eds.), *Physical Acoustics*, Vol. XIX, *Ultrasonic Measurement Methods*, Academic, Boston, 1990, pp. 81–106 and 107–155.
12. D. L. Arenburg, *J. Acoust. Soc. Am.*, Vol. 20, No. 1, 1948, pp. 1–26.
13. P. M. Morse and K. U. Ingard, *Theoretical Acoustics*, McGraw-Hill, New York, 1968, pp. 400–441.
14. O. E. Mattiat, *Ultrasonic Transducer Materials*, Plenum, New York, 1971.
15. G. Simmons and H. Wang, *Single Crystal Elastic Constants and Calculated Aggregate Properties: A Handbook*, MIT Press, Cambridge, MA, 1971.
16. L. M. Brekhovskikh, *Waves in Layered Media*, 2nd ed., Academic, New York, 1980, pp. 14, 38–41, 48–50, 53, 95.
17. L. M. Brekhovskikh, *Waves in Layered Media*, 2nd ed., Academic Press, New York, 1980, pp. 60–70. See also I. A. Vicktorov, *Rayleigh and Lamb Waves*, Plenum, New York, 1967.
18. T. R. Meeker and A. H. Meitzler, "Guided Wave Propagation in Elongated Cylinders and Plates," in W. P. Mason (Ed.), *Physical Acoustics: Principles and Methods*, Vol. I A, Academic, New York, 1964, pp. 111–167.
19. T. R. Meeker and A. H. Meitzler, "Guided Wave Propagation in Elongated Cylinders and Plates," in W. P. Mason (Ed.), *Physical Acoustics: Principles and Methods*, Academic, New York, 1964, especially pp. 134–140.
20. R. C. McMaster (Ed.), *Nondestructive Testing Handbook*, Vol. II, American Society for Nondestructive Testing; Roland, New York, 1959, pp. 43.8–43.11.

21. E. P. Papadakis, "Ultrasonic Velocity and Attenuation: Measurement Methods with Scientific and Industrial Applications," in W. P. Mason and R. N. Thurston (Eds.), *Physical Acoustics: Principles and Methods*, Vol. XII, Academic, New York, 1976, pp. 277–374.
22. E. P. Papadakis, "Physical Acoustics and the Microstructure of Iron Alloys," *Inter. Met. Rev.*, Vol. 29, No. 1, 1984, pp. 1–24.
23. E. P. Papadakis, "A Computer-Automated Statistical Process Control Method with Timely Response," *Eng. Costs Product Econ.*, Vol. 18, 1990, pp. 301–310.
24. E. P. Papadakis, "Sampling Plans and 100% Nondestructive Inspection Compared," *Quality Progress*, Vol. 15, No. 4, 1982, pp. 37–39.
25. E. P. Papadakis, "Ultrasonic Attenuation Caused by Scattering in Polycrystalline Media," in W. P. Mason (Ed.), *Physical Acoustics: Principles and Methods*, Vol. IVB, Academic, New York, 1968.
26. H. Seki, A. Granato, and R. Truell, "Diffraction Effects in the Ultrasonic Field of a Piston Source and Their Importance in the Accurate Measurement of Attenuation," *J. Acoust. Soc. Am.*, Vol. 28, No. 2, 1956, pp. 230–238.
27. E. P. Papadakis, "Ultrasonic Attenuation and Velocity in Three Transformation Products in Steel," *J. Appl. Phys.*, Vol. 35, 1964, pp. 1474–1482.
28. W. P. Mason, *Physical Acoustics and the Properties of Solids*, Van Nostrand, Princeton, 1958, p. 27.
29. M. G. Silk, *Ultrasonic Transducers for Nondestructive Testing*, Adam Hilger, Bristol, England, 1984.
30. E. P. Papadakis, "Theoretical and Experimental Methods to Evaluate Ultrasonic Transducers for Inspection and Diagnostic Applications," *IEEE Trans.*, Vol. SU-26, 1979, pp. 14–27.
31. W. P. Mason, *Physical Acoustics and the Properties of Solids*, Van Nostrand, Princeton, 1958, pp. 23–32.
32. E. P. Papadakis, "Lens Equation for Focused Transducers," *Int. J. Nondestruct. Test.*, Vol. 4, 1972, pp. 195–198.
33. R. R. Price and A. C. Fleischer, "Sonographic Instrumentation," in A. C. Fleischer et al., *Ultrasonography in Obstetrics and Gynecology*, 4th ed., Appleton & Lange, Norwalk, CT, 1991, pp. 25–38.
34. T. M. Mansour, "Evaluation of Ultrasonic Transducers by Cross-Sectional Mapping of the Near Field Using a Point Reflector," *Mat. Evaluation*, Vol. 37, No. 7, 1977, pp. 50–54.
35. R. Straw, "Do You Hear What I Hear?—The Early Days of Pulse-Echo Ultrasonics," *Mat. Evaluation*, Vol. 42, No. 1, 1984, pp. 24–28.
36. F. A. Firestone, and J. R. Fredericks, "Refinements in Supersonic Reflectoscopy. Polarized Sound," *J. Acoust. Soc. Am.*, Vol. 18, No. 1, 1945, pp. 200–211.
37. E. P. Papadakis, "Broadband Flaw Detection Transducers: Application to Acoustic Emission Pulse Shape and Spectrum Recording Based on Pulse Echo Response Spectrum Corrected for Beam Spreading," *Acustica*, Vol. 46, 1980, pp. 293–298.
38. W. Woodmansee and R. D. Whealy, Boeing Commercial Airplane Division, Seattle, Washington, private communication.
39. T. Sato and O. Ikeda, "Sequential Synthetic Aperture Sonar System: A Prototype of a Synthetic Aperture Sonar System," *IEEE Trans.*, Vol. SU-24, 1977, pp. 253–259.
40. R. K. Muller, M. Kaveh, and G. Wade, "Reconstructive Tomography and Applications to Ultrasound," *IEEE Proc.*, Vol. 67, 1979, pp. 567–587.
41. R. Truell, C. Elbaum, and B. B. Chick, *Ultrasonic Methods in Solid State Physics*, Academic Press, New York, 1969.
42. E. P. Papadakis, "New, Compact Instrument for Pulse-Echo-Overlap Measurements of Ultrasonic Wave Transit Times," *Rev. Sci. Instrum.*, Vol. 47, 1976, pp. 806–813.
43. E. P. Papadakis, "Absolute Accuracy of the Pulse-Echo-Overlap Method and the Pulse-Superposition Method for Ultrasonic Velocity," *J. Acoust. Soc. Am.*, Vol. 52 (Part 2), 1972, pp. 843–846.
44. J. E. May, Jr., "Precise Measurement of Time Delay," *IRE Ntl. Convection Rec.*, Vol. 6 (Part 2), 1958, pp. 134–142.
45. E. P. Papadakis, "Ultrasonic Phase Velocity by the Pulse-Echo-Overlap Method Incorporating Diffraction Phase Corrections," *J. Acoust. Soc. Am.*, Vol. 42, 1967, pp. 1045–1051.
46. E. P. Papadakis, "Variability of Ultrasonic Shear Wave Velocity in Vitreous Silica for Delay Lines," *IEEE Trans.*, Vol. SU-16, 1969, pp. 210–218.
47. G. A. Coquin and R. Tsu, "Theory and Performance of Perpendicular Diffraction Delay Lines," *Proc. IEEE*, Vol. 53, No. 6, 1965, pp. 581–591.
48. E. Plenard, "The Elastic Behavior of Cast Iron," National Metal Congress, Cleveland, 1964.
49. Engineering and Research Staff, *Potential Failure Mode and Effects Analysis: An Instruction Manual*, Ford Motor Company, Dearborn, MI, September, 1979.
50. "Eighth Annual NDT Buyers' Guide," *Mat. Evaluation*, Vol. 48, No. 6, 1990, pp. 725–784.

61

THERMOACOUSTIC ENGINES

GREGORY W. SWIFT

1 INTRODUCTION

Thermoacoustic engines harness the thermal effects present in sound waves to produce acoustic power from heat or to produce refrigeration from acoustic power. In most acoustic phenomena, the temperature oscillation accompanying the pressure oscillation in a sound wave is unimportant, except as it contributes to the attenuation of sound in bulk and at boundaries. However, under some circumstances the temperature oscillation and in particular its interaction with boundaries—thermoacoustic processes—can be harnessed to produce powerful, useful thermodynamic effects. Both basic types of heat engine can be realized thermoacoustically: In a thermoacoustic prime mover, heat flow from high temperature to lower temperature produces work in the form of acoustic power; in a thermoacoustic heat pump, acoustic power is used to pump heat from low temperature to higher temperature. Either a predominantly standing wave or a predominantly traveling wave can be used, although geometries for the two types of waves are substantially different. Qualitative understanding of thermoacoustic engines is easiest when considering the position, pressure, volume, and temperature oscillations of typical parcels of gas. Quantitative analysis is based on the acoustic approximation.

2 EXAMPLES

The earliest thermoacoustic engine was the Sondhauss tube, shown in Fig. 1*a*. Over 100 years ago, glassblowers found that when a small, hot glass bulb was being blown on a cool glass tubular stem, the stem tip sometimes radiated sound. Rayleigh[1] understood that the sound was generated by oscillatory thermal expansion and contraction of the air in the tube, which in turn was due to acoustic motion of the air toward and away from the hot end of the tube. The Sondhauss tube is a thermoacoustic prime mover, converting heat into mechanical work in the form of sound. It utilizes a predominantly standing acoustic wave, with the bulb and stem forming a resonator. A key geometric requirement for this and other standing-wave thermoacoustic engines is that, in the region of substantial temperature gradient, the smallest transverse dimension must be comparable to the thermal penetration depth in the air, which is roughly the distance that heat can diffuse in a time $1/\pi f$, where f is the frequency of the wave.

Hofler's thermoacoustic refrigerator[2] is another example of a standing-wave thermoacoustic engine and illustrates the separate functions of resonator, power transducer, and heat exchanger components. As shown in Fig. 1*b*, it consisted of a helium-filled resonator driven by a loudspeaker and containing a "stack" with one end thermally anchored at room temperature by a hot heat exchanger. With a static pressure of 1 MPa and driven at an amplitude of 30 kPa, its cold heat exchanger reached 200 K; at somewhat higher temperatures, it produced 3 W of cooling power with a coefficient of performance of 0.12 of the Carnot coefficient of performance.

The resonator, consisting of tubing segments and a spherical bulb, produced a resonance frequency of 500 Hz compactly and with minimal dissipation of acoustic power. The heart of this thermoacoustic refrigerator was the stack, a ribbon of plastic sheet spirally wound to make a 3.8-cm-diameter, 8-cm-long assembly. The spacing of 0.4 mm between turns of the spiral was maintained by plastic spacers; this spacing is several thermal penetration depths at the 500-Hz operating frequency and

---

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker
Copyright © 1997 John Wiley & Sons, Inc.

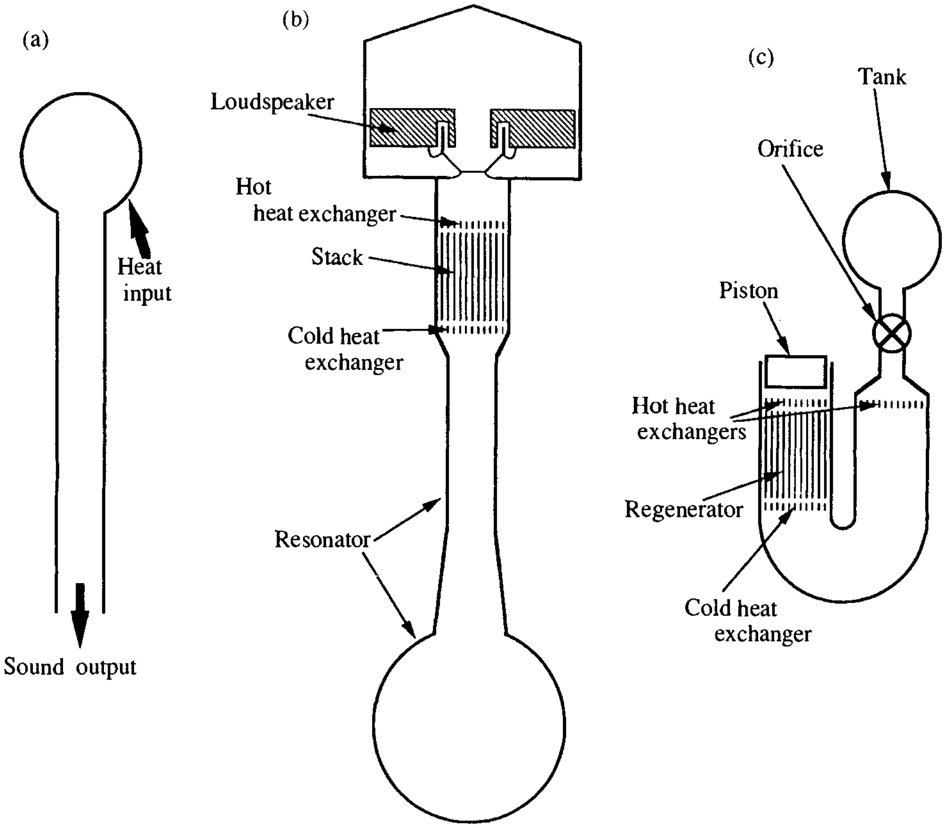


Fig. 1 Examples of thermoacoustic engines: (*a*) the Sondhauss tube, a standing-wave thermoacoustic prime mover; (*b*) Hofler's refrigerator, a standing-wave thermoacoustic heat pump; (*c*) the orifice pulse-tube refrigerator, a traveling-wave thermoacoustic heat pump.

enabled heat to be pumped thermoacoustically from the cold heat exchanger to the hot heat exchanger. These heat exchangers consisted of well-spaced copper strips extending across the resonator at either end of the stack to ensure thermal contact between the gas at the stack ends and the resonator walls nearby.

Ceperley[3] has discussed thermoacoustic engines using traveling waves. One example is the orifice pulse-tube refrigerator, reviewed by Radebaugh[4] and shown in Fig. 1*c*. A piston functions as a source of acoustic power at one end of the refrigerator, and an orifice and tank act as a sink of acoustic power at the other end. Between them, a regenerator, heat exchangers, and an open tube are filled with gas. The regenerator and adjacent heat exchangers function just like those of a Stirling cycle refrigerator and are the heart of this refrigerator; with the hot heat exchangers anchored at room temperature, the cold heat exchanger reaches typically 40 K, and refrigeration is produced with a coefficient of performance of typically one-fourth of Carnot's. The key geometric requirement is that the smallest transverse dimension in the regenerator be much smaller than the thermal penetration depth in the gas at the operating frequency; this requirement is characteristic of traveling-wave thermoacoustic engines, in contrast to standing-wave thermoacoustic engines. The open tube functions as a temperature buffer space between the cold heat exchanger and the warm orifice.

Many attempts to produce practical and commercially viable thermoacoustic devices are underway worldwide. Standing-wave refrigerators having cooling powers of tens or hundreds of watts at 0°C are under development for satellite and residential use. The technology is inexpensive, reliable, and environmentally benign; commercial success will require improvements in efficiency. Standing-wave prime movers driving orifice pulse tube refrigerators comprise the only cryogenic refrigeration technology having no moving parts; this combination is under development for cooling cryogenic electronics and sensors and for cryogen liquefaction. Again, low cost and reliability are the technology's chief virtues; developers are striving to increase efficiency.

3 LAGRANGIAN EXPLANATION OF PHENOMENA

Time phasing is an important factor in the operation of traditional heat engines: Pistons, valves, and sparks must have the correct relative timing in order to produce the desired thermodynamic cycle in the working gas. Thermoacoustic engines contain no obvious moving parts to perform these functions, yet the acoustic stimulation of heat flux and the generation (or absorption) of acoustic power point to some type of timed phasing of thermodynamic processes, here achieved in a remarkably simple way.

The key to phasing in thermoacoustic engines is the presence of two thermodynamic media: gas and solid. As the gas oscillates along the solid at the acoustic frequency, it experiences oscillations in temperature. Part of the temperature oscillations comes from compression and expansion of the gas by sound pressure, and the rest is a consequence of the local temperature of the solid itself.

To understand thermoacoustic phenomena qualitatively, it is best to adopt a Lagrangian point of view, following a typical parcel of gas as it moves. Real thermoacoustic engines operate sinusoidally, but it is simplest here to consider square-wave motion and pressure in the gas, with pressure and displacement in phase for standing waves and pressure and velocity in phase for traveling waves. Viscosity is ignored in this section, even though it has a significant, harmful effect on thermoacoustic efficiency. With these conceptual simplifications, the thermodynamic cycle for standing waves becomes identical to the Brayton cycle, with two reversible adiabats and two irreversible constant-pressure heat transfers, while the cycle for traveling waves becomes two isotherms and two reversible polytropic processes.

3.1 Standing-Wave Heat Pump

The standing-wave thermoacoustic heat pump can be understood with reference to Fig. 2, where the four steps of its cycle are displayed for one typical parcel of gas. The gas oscillates along the adjacent solid in steps 1 and 3 in response to the standing wave and is also adiabatically pressurized and depressurized in steps 1 and 3, respectively, absorbing and doing work dW''. Steps 2 and 4 allow thermal equilibration between gas and adjacent solid.

There is a critical longitudinal temperature gradient $\nabla T_{\text{crit}} = p/\xi\rho c_p$ at which no heat is transferred in steps 2 and 4. (Here, p is the acoustic pressure amplitude, ξ is the acoustic displacement amplitude, ρ is density, and c_p is isobaric heat capacity per unit mass.) This occurs when the temperature rise due to adiabatic compression in step 1, $2p/\rho c_p$, exactly matches the rise in adjacent wall temperature that the gas experiences because of its motion along the wall $2\xi\,\nabla T_{\text{crit}}$. For all temperature gradients less than ∇T_{crit}, heat transfer between gas and solid is as shown in Fig. 2, with heat flow dQ from gas to solid in step 2, at the higher temperature and position, and heat flow dQ' from solid to gas in step 4, at the lower temperature and position. Thus, the net effect of this parcel of gas is to move a little heat a short distance up the solid, from slightly lower temperature to slightly higher temperature, during each cycle of the sound wave. All the other parcels do the same thing; so the overall effect,

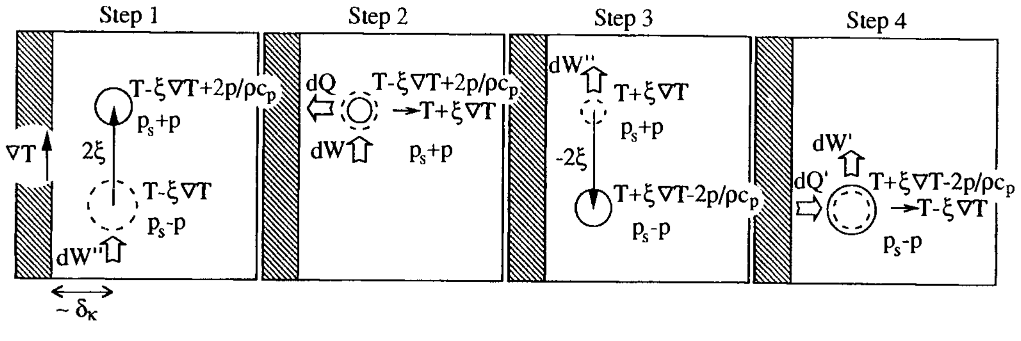


Fig. 2 Typical gas parcel executing the four steps of the cycle of a standing-wave thermoacoustic heat pump, assuming an inviscid gas and square-wave acoustic motion and pressure. In each step, the dashed and solid circles show respectively the initial and final positions and volumes of the parcel, while labels show the initial and final temperatures and pressures of that step, and broad arrows show directions of attendant heat flows (either from the parcel to the solid or from the solid to the parcel) and work flows (either into or out of the parcel). Averaged over one cycle, the parcel absorbs work $dW - dW'$ from the wave.

much as in a bucket brigade, is the net transport of heat from the cold end of the stack to its hot end.

Imperfect thermal contact between gas and solid is required for the cycle to operate as described. If thermal contact between gas and solid were perfect, the gas would arrive at its uppermost position at the end of step 1 already equilibrated at the local solid temperature, so that no heat would be transferred during step 2; and similarly for steps 3 and 4. At the other extreme, if thermal contact were nonexistent, no heat could be transferred in steps 2 and 4. Thus, intermediate, *imperfect* thermal contact between gas and solid is required for the cycle to operate as described. For the sinusoidal motion and pressure characteristic of real engines, this imperfect thermal contact is achieved for gas neither much closer nor much farther than a thermal penetration depth from the nearest solid. Thus, effective standing-wave engines fill the entire cross section of the resonator with a stack having spacings of a few thermal penetration depths, as shown in Fig. 1*b*.

3.2 Standing-Wave Prime Mover

Similarly, the standing-wave thermoacoustic prime mover is illustrated in Fig. 3. Here, the temperature gradient is greater than ∇T_{crit}, so the directions of heat transfer in steps 2 and 4 are reversed relative to the standing-wave heat pump. Thus, the net thermal effect of the parcel of gas shown is to move a little heat down the solid, from higher to lower temperature. Again, all parcels have a similar effect, so that overall a substantial amount of heat is carried down along the solid from the hot heat exchanger to the cold heat exchanger.

Since heat flows into the gas in step 2 and out of it in step 4, the gas expands at high pressure in step 2 and contracts at low pressure in step 4. Thus, the net effect in the thermoacoustic prime mover is the production of work at the acoustic frequency by this and all other gas parcels. One can imagine all the gas in the stack region of a thermoacoustic prime mover thermally expanding and contracting because of its displacement along the temperature gradient, with the correct phasing relative to the oscillatory standing-wave pressure to do net work. In the Sondhauss tube, for example, this net work maintains the standing wave against viscous and thermal losses in the resonator and provides power to radiate into the room.

3.3 Traveling-Wave Heat Pump

Traveling-wave thermoacoustic engines differ from standing-wave engines in two fundamental ways: phasing between pressure and motion and spacing between solid elements. These differences are illustrated for the traveling-wave heat pump in Fig. 4. Isothermal compression in step 2 transfers heat from the gas parcel to the solid at a higher temperature, and isothermal expansion at step 4 transfers heat from solid to gas at a lower temperature. The other two steps serve to preheat and precool the gas, reversibly storing and recovering heat in the solid. The net effect of one parcel over an entire cycle is, again, to move a little heat a short distance up the solid, up the temperature gradient. Excellent thermal contact between gas and solid is required in steps 1 and 3 for effective preheating and precooling, respectively; in realistic, sinusoidal traveling-wave engines, this is achieved by maintaining transverse dimensions much smaller than a thermal penetration depth.

Traveling-wave prime movers are also possible in principle,[3] although none has been built.

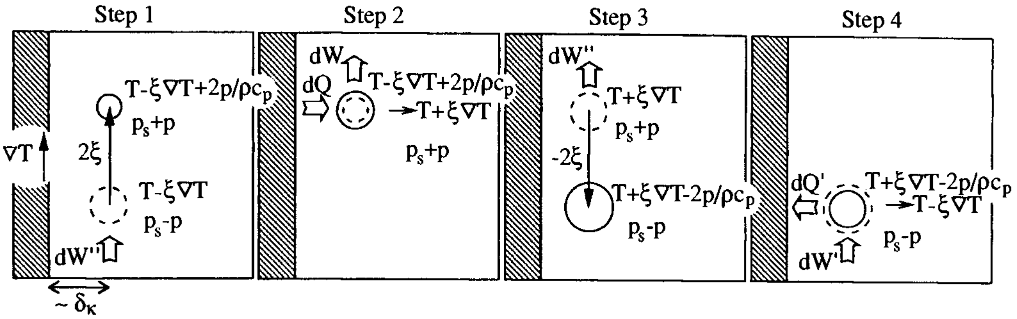


Fig. 3 Typical gas parcel executing the four steps of the cycle of a standing-wave thermoacoustic prime mover, assuming an inviscid gas and square-wave acoustic motion and pressure. Symbols have the same meaning as in Fig. 2. Averaged over one cycle, the parcel delivers work $dW - dW'$ to the wave.

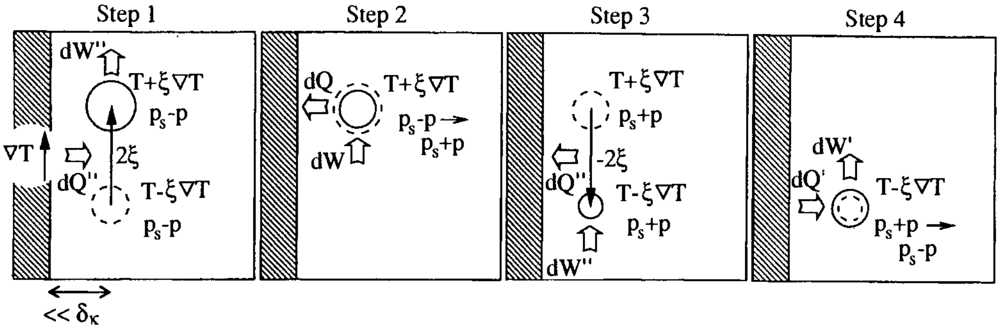


Fig. 4 Typical gas parcel executing the four steps of the cycle of a traveling-wave thermoacoustic heat pump, assuming an inviscid gas and square-wave acoustic motion and pressure. Symbols have the same meaning as in Fig. 2.

4 METHODS OF COMPUTATION

4.1 Accurate Method

Applicable to both standing- and traveling-wave thermoacoustic engines, the wave equation for the complex pressure amplitude p in a gas in a small channel with a temperature gradient is[5]

$$\frac{\omega^2}{c^2}[1+(\gamma-1)f_\kappa]p+\rho\frac{d}{dx}\left[\frac{(1-f_\nu)}{\rho}\frac{dp}{dx}\right] - \frac{1}{T}\frac{dT}{dx}\frac{f_\kappa-f_\nu}{1-\sigma}\frac{dp}{dx}=0, \tag{1}$$

where f_ν and f_κ are defined below. The acoustic power flowing through the channel is

$$W=\frac{A}{2\omega\rho}\,\mathrm{Im}\left[p\,\frac{d\tilde{p}}{dx}\,(1-\tilde{f}_\nu)\right] \tag{2}$$

and the enthalpy flow (total energy flow) through the channel is

$$H=\frac{A}{2\omega\rho}\,\mathrm{Im}\left[p\,\frac{d\tilde{p}}{dx}\left(1-\tilde{f}_\nu-\frac{f_\kappa-\tilde{f}_\nu}{1+\sigma}\right)\right] + \frac{Ac_p}{2\omega^3\rho(1-\sigma)}\frac{dT}{dx}\frac{dp}{dx}\frac{d\tilde{p}}{dx}\,\mathrm{Im}\left(\tilde{f}_\nu+\frac{f_\kappa-\tilde{f}_\nu}{1+\sigma}\right) - (AK+A_sK_s)\frac{dT}{dx}. \tag{3}$$

Here

$$f_\nu=\frac{\tanh[(1+i)y_0/\delta_\nu]}{(1+i)y_0/\delta_\nu},\qquad f_\kappa=\frac{\tanh[(1+i)y_0/\delta_\kappa]}{(1+i)y_0/\delta_\kappa} \tag{4}$$

for parallel-plate channels of separation $2y_0$; $\delta_\nu=\sqrt{2\nu/\omega}=\sqrt{2\mu/\omega\rho}$ and $\delta_\kappa=\sqrt{2\kappa/\omega}=\sqrt{2K/\omega\rho c_p}$ are the viscous and thermal penetration depths, where ν is kinematic viscosity, μ is dynamic viscosity, κ is thermal diffusivity, and $\omega=2\pi f$ is the radian frequency of the wave; σ is the Prandtl number, γ is the ratio of isobaric to isochoric specific heats, and T is the temperature; A and A_s and K and K_s are the areas and thermal conductivities of gas and solid, respectively; and $i=\sqrt{-1}$, Im signifies the imaginary part of what follows, and the tilde denotes complex conjugation. Results shown here are for ideal gases only; results for liquids can be found in Ref. 6. Expressions for f_ν and f_κ for circular channels, rectangular and triangular channels, and pin arrays can be found in Refs, 5, 7, and 8, respectively.

When numerically integrating these equations to predict the performance of thermoacoustic engines, recognize that Eqs. (1) and (3) comprise a set of three second-order coupled real differential equations in the variables $T(x)$, Re[$p(x)$], and Im[$p(x)$]. With the addition of the definition

$$p_x=\frac{dp}{dx}, \tag{5}$$

Eqs. (1), (3), and (5) become a set of five first-order coupled real differential equations in x in the five variables T, Re[p], Im[p], Re[p_x], and Im[p_x]. The enthalpy flow H is a given constant, independent of x within the stack or regenerator because a steady state is assumed and no energy flows in or out in the transverse directions (except at the heat exchangers, where the value of H is established). The gas and solid thermophysical properties depend implicitly on x through their dependence on T.

A solution may be obtained numerically using a Runge–Kutta method starting at one end of the stack or regenerator and integrating to the other end. Required to start the integration is a choice of five attendant initial conditions: the temperature, complex pressure amplitude, and complex impedance at the starting position.

Such a Runge–Kutta thermoacoustic code has been imbedded as a subroutine in a more complete program[9] including resonator shape and loss, power transducer impedance, heat exchanger geometry, and the like. The larger program can iteratively call the thermoacoustic subroutine, so that the user can supply more convenient boundary conditions such as hot and cold temperatures, and the program can then find variables such as resonance frequency, acoustic power, and enthalpy flow. The results are generally in good agreement with experimental data.[10]

4.2 Approximate Method, Standing Waves

The energy flows in parallel-plate standing-wave thermoacoustic engines are given approximately[6] by

$$
\begin{aligned}
H \simeq{}& \frac{A\delta_\kappa}{4y_0} \frac{p\langle u\rangle}{(1+\sigma)(1-\delta_\nu/y_0+\delta_\nu^2/2y_0^2)} \\
&\cdot \left(\Gamma \frac{1+\sqrt{\sigma}+\sigma}{1+\sqrt{\sigma}} - 1 - \sqrt{\sigma} + \frac{\delta_\nu}{y_0}\right) \\
&+ (AK + A_sK_s)\,\frac{\Delta T}{\Delta x},
\end{aligned} \tag{6}
$$

$$
\begin{aligned}
W \simeq{}& \frac{A\delta_\kappa\,\Delta x(\gamma-1)\omega p^2}{4y_0\rho c^2} \\
&\cdot \left[\frac{\Gamma}{(1+\sqrt{\sigma})(1-\delta_\nu/y_0+\delta_\nu^2/2y_0^2)} - 1\right] \\
&- \frac{A\delta_\nu\,\Delta x}{4y_0} \frac{\omega\rho\langle u\rangle^2}{(1-\delta_\nu/y_0+\delta_\nu^2/2y_0^2)},
\end{aligned} \tag{7}
$$

where

$$
\Gamma = \frac{\rho c_p\langle u\rangle}{\omega p} \frac{\Delta T}{\Delta x}, \tag{8}
$$

Δx is the length of the engine along the propagation direction, ΔT is the temperature difference spanned, and $\langle u\rangle$ is the transverse spatial average (perpendicular to the gas motion) of the velocity amplitude u. Here, the sign conventions are that ΔT, p, and $\langle u\rangle$ are all real and positive and H (the enthalpy flow) and W (the net acoustic power) are positive for a prime mover. It is assumed that the engine is short enough that thermophysical properties p and u are essentially independent of x and can be approximated by their values in the middle of the engine and further that the channels are large enough (typically $y_0 \geq 2\delta_\kappa$) that the hyperbolic tangents in Eqs. (4) are essentially unity. Equations (6) and (7) are generally accurate to within a factor of 2. They are easily computed with a programmable calculator or simple computer program and are useful in finding overall trends and approximate magnitudes.

For $\Gamma = 0$ and $\delta_\nu/y_0 \ll 1$, Eq. (7) reduces to the expression for thermoviscous wall losses in standing waves (cf. Chapter 22).

4.3 Approximate Method, Traveling Waves

The order of magnitude of the gross refrigeration power in well-designed traveling-wave thermoacoustic heat pumps is[4]

$$
H \simeq \frac{Ap\langle u\rangle T_C}{30T_H}, \tag{9}
$$

where T_C and T_H are the cold and hot heat exchanger temperatures. Considering only regenerator effects, the efficiency is approximately half of Carnot's efficiency.

5 AUXILIARY COMPONENTS

Two heat exchangers are required in thermoacoustic engines to supply and extract heat at the two ends of the stack or regenerator. The length of each along the direction of the acoustic oscillation should be of the order of the local displacement amplitude ξ; transverse dimensions should be of the order of those of the adjacent stack or regenerator. Equation (7) with $\Gamma = 0$ can be used to estimate the acoustic power dissipation in the heat exchangers.

The heat exchangers for small engines can simply consist of multiple copper strips spanning the engine at the stack ends. Larger diameter engines may require integral heat pipes, pumped-fluid channels, or electric resistance heaters imbedded in the heat exchangers.

A resonator is required for standing-wave thermoacoustic engines. It should minimize unnecessary dissipation of acoustic power into heat. Unusual geometries[2] such as that of Fig. 1*b* can reduce resonator dissipation by almost a factor of 2 relative to uniform-diameter right-circular-cylinder resonators. Resonator dissipation can also be calculated using Eq. (7) with $\Gamma = 0$.

6 OVERALL PERFORMANCE

Calculations using Eqs. (1) and (3) generally show that standing-wave thermoacoustic engines have second-law efficiencies (i.e., efficiencies normalized by the Carnot efficiency) between 0.2 and 0.4, considering effects in the stack only. Including resonator, heat exchanger, and transducer losses reduces this efficiency to between 0.15 and 0.3 for large-cross-section engines, for which resonator losses are not a significant fraction of the total power; small-diameter engines have even lower efficiencies. Traveling-wave thermoacoustic engines can have higher efficiencies.

The power per unit of cross-sectional area in standing-wave thermoacoustic engines can be estimated crudely by $H/A \sim (p/p_s)^2 p_s c/10$. Thus, high amplitude, high static pressure, and high sound speed increase power density. Heat flows of the order of 1 MW/m^2 are easily achievable in high-pressure helium gas. The pressure amplitude is often 5–10% of static pressure; the gas displacement amplitude can exceed the length of heat exchangers and approach the length of the stack. At such high amplitudes, performance deviates significantly[10] from the predictions of Eqs. (1) and (3).

REFERENCES

1. Lord Rayleigh (J. W. Strutt), *The Theory of Sound*, 2nd ed., Vol. 2, Dover, New York, 1945, Sec. 322.
2. T. Hofler, "Thermoacoustic Refrigerator Design and Performance," Ph.D. Dissertation, University of California at San Diego, 1986; summarized in T. Hofler, "Concepts for Thermoacoustic Refrigeration and a Practical Device," in *Proc. 5th Int. Cryocoolers Conf.*, August 1988, Monterey CA, p. 93.
3. P. H. Ceperley, "Gain and Efficiency of a Short Traveling Wave Heat Engine," *J. Acoust. Soc. Am.*, Vol. 77, 1985, p. 1239.
4. R. Radebaugh, "A Review of Pulse Tube Refrigeration," in *Adv. Cryogenic Eng.*, Vol. 35, Plenum, New York, 1990, p. 1191.
5. N. Rott, "Damped and Thermally Driven Acoustic Oscillations in Wide and Narrow Tubes," *Z. Angew. Math. Phys.*, Vol. 20, 1969, p. 230; "Thermally Driven Acoustic Oscillations, Part III: Second-Order Heat Flux," *Z. Angew. Math. Phys.*, Vol. 26, 1975, p. 43.
6. G. W. Swift, "Thermoacoustic Engines," *J. Acoust. Soc. Am.*, Vol. 84, 1988, p. 1145.
7. W. P. Arnott, H. E. Bass, and R. Raspet, "General Formulation of Thermoacoustics for Stacks Having Arbitrarily Shaped Pore Cross Sections," *J. Acoust. Soc. Am.*, Vol. 90, 1991, p. 3228.
8. G. W. Swift and R. M. Keolian, "Thermoacoustics in Pin-Array Stacks," *J. Acoust. Soc. Am.*, Vol. 94, 1993, p. 941.
9. W. C. Ward and G. W. Swift, "Design Environment for Low Amplitude Thermoacoustic Engines," *J. Acoust. Soc. Am.*, Vol. 95, 1994, p. 3671. Fully tested software and users guide available from Energy Science and Technology Software Center, US Department of Energy, Oak Ridge TN. For a beta-test version, contact ww@lanl.gov (Bill Ward) via Internet.
10. J. R. Olson and G. W. Swift, "Similitude in Thermoacoustics," *J. Acoust. Soc. Am.*, Vol. 95, 1994, p. 1405.

PART VI

MECHANICAL VIBRATIONS AND SHOCK

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker
Copyright © 1997 John Wiley & Sons, Inc.

62

INTRODUCTION

MANFRED A. HECKL

1 INTRODUCTION

Within the scope of acoustics mechanical vibrations are defined as motions of rigid or elastic solid bodies provided the spectrum of the motion contains frequencies within the acoustic range. Other names for this field are structure-borne sound or vibroacoustics. Typical topics of mechanical vibrations in acoustics are as follows:

- Structural vibrations caused by unbalanced forces, misalignments, rolling over rough surfaces, internal combustion, parametric excitation, stick-slip, and so on. Such phenomena occur in all kinds of machines and vehicles. They are of great importance for noise control problems (see also Chapter 71).
- Vibrations of space vehicles (acoustic fatigue), aircraft surfaces, windows, walls and so on, caused by strong airborne sound sources or turbulent flows (see Chapters 11 and 70).
- Sound transmission through walls, ceilings, enclosures.
- Shock and vibration isolation (see Chapter 67).
- Ground vibrations near factories, forge hammers, underground railways, and so on.
- Measurement of elastic moduli and loss factors of materials (mechanical spectroscopy) (see Chapter 72).
- High-frequency elastic waves in delay lines or in (ultrasonic) nondestructive testing equipment (see Chapters 58 and 69)
- Machine monitoring and machine diagnosis (see Chapters 68 and 73).

Shocks constitute a subclass of mechanical vibrations. They are associated with rather short and sudden signals. Usually they are of high amplitude; therefore, nonlinear effects are of concern for shock absorbers. Consideration of shocks is important in packaging, in the field of machine reliability, and in military applications (see Chapter 67).

2 BASIC RELATIONS

2.1 Newton's Law and Hooke's Law

When the vibrations are small enough so that the principle of superposition (linearity) can be employed and when losses are negligible, all equations dealing with mechanical vibrations can be seen as a consequence of Newton's law and Hooke's law. For lumped-parameter systems consisting of masses m and massless springs of stiffness s, these laws are

$$m\frac{\partial^2 \xi_i}{\partial t^2} = m\frac{\partial v_i}{\partial t} = F_i \quad \text{(Newton)},$$

$$F_i = s_i u_i \quad \text{(Hooke)}. \qquad (1)$$

These two equations describe in most simple terms the inertia forces and the elastic forces necessary for the existence of elastic waves. The terms ξ_j and v_i are the components of displacement and velocity of the mass; F_i are the components of the force acting on m; and u_i is the length change of the spring due to the force F_i.

The equations contain two approximations. One of them consists in the replacement of a total derivative by a partial derivative; that is, convective terms and very high amplitudes are excluded. The second is that the spring behaves linearly.

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker

Copyright © 1997 John Wiley & Sons, Inc.

In Chapter 63 Newton's and Hooke's laws are applied extensively to derive the equations of motion of single- and multi-degree-of-freedom mechanical systems. If the two laws are applied to continua, the mass has to be replaced by a mass per unit volume (i.e., density ρ), the forces become stresses σ_{ij}, and the relative length change has to be replaced by the strains. This way one finds (see, e.g., Refs. 1 and 2)

$$\rho \frac{\partial^2 \xi_i}{\partial t^2} = \frac{\partial \sigma_{ij}}{\partial x_j} + F_i'',$$

$$\sigma_{ij} = C_{ijkl} \frac{1}{2} \left(\frac{\partial \xi_k}{\partial x_l} + \frac{\partial \xi_l}{\partial x_k} \right). \qquad (2)$$

Here F_i'' is the force (if any) per unit volume that acts from outside and C_{ijkl} is the elasticity tensor characteristic of the material. In these expressions the summation convention has been used; that is, one has to sum over identical indices.

For the most important case, a homogeneous, isotropic material with shear modulus G and Poisson number μ the stresses in Eq. (2) can be written as

$$\sigma_{ij} = G \left[\frac{2\mu}{1 - 2\mu} \left(\frac{\partial \xi_k}{\partial x_k} \right) \delta_{ij} + \left(\frac{\partial \xi_i}{\partial x_j} + \frac{\partial \xi_j}{\partial x_i} \right) \right]. \qquad (3)$$

Instead of G and μ one might write Eq. (3) also in terms of G and Young's modulus E, or (as in many books on theoretical physics) in terms of the so-called Lamé constants λ and μ_L. These quantities are related to each other by

$$E = 2(1 + \mu)G, \qquad \lambda = \frac{E\mu}{(1 + \mu)(1 - 2\mu)}, \qquad \mu_L = G. \qquad (4)$$

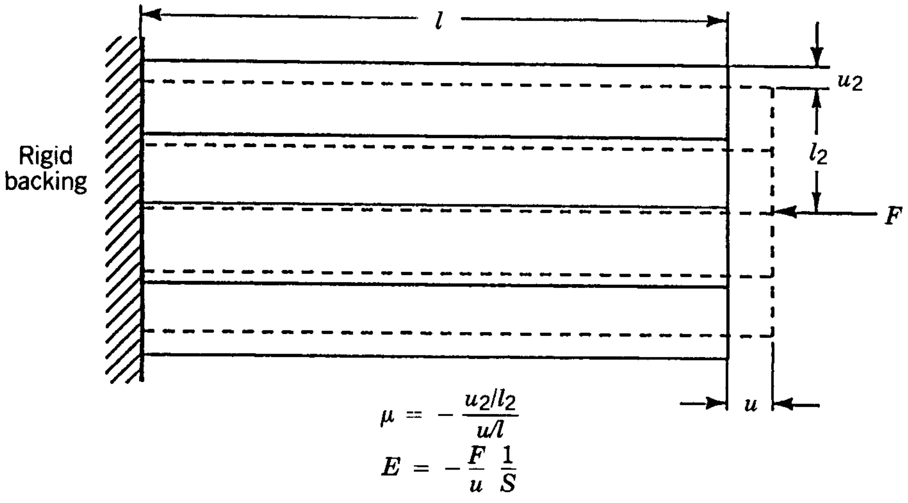
For short bars of cross section S and length l Young's modulus E and μ are defined in Fig. 1. As can be seen, μ is related to lateral contraction or extension when a bar is elongated or compressed.

The main difference between waves in gases or fluids and in elastic solids is that, in the absence of loss mechanisms, there are only compressional (longitudinal) waves in gases or fluids, whereas there can be compressional and shear waves in solids. Other wave types such as bending waves, torsional waves, and Rayleigh (surface) waves can be understood as superpositions of compressional and shear waves.

2.2 Loss Mechanisms

In the absence of loss mechanisms or damping the vibratory energy in finite systems would remain constant if after some sort of excitation it is left free. As a consequence, a system would never "forget" its past history; for example, a simple resonator that is driven since $t = 0$ with a frequency close to its resonance frequency would exhibit the beating frequency forever. Since such a behavior does not agree with reality, the basic equations have to be modified to account for the effect of losses.

The usual way to include linear damping in the basic equations is to modify the force–displacement or



$S = l_2 l_3$ = cross section, l_3 = width

Fig. 1 Definition of Poisson's ratio and Young's modulus for static loads: (----) undeflected state; (——) compressed state.

stress–strain relation in Eqs. (1) and (2). In the simplest case a velocity-proportional, viscous force is added so that the second part of Eqs. (1) becomes

$$F = su + b\,\frac{\partial u}{\partial t} = su + bv. \tag{5}$$

Here b is the damping constant and v the relative velocity of the two sides of the spring. For elastic continua a viscous term has to be added to Eq. (2). Normally this is not done in acoustics. The concept of complex moduli [see Fig. (8)] is employed instead.

Equation (5) is suitable to represent dampers that are in parallel to springs, but they do not correctly describe the behavior of materials with inner losses. Other and, if the parameters are chosen properly, better relations for lossy, elastic structural elements are[3]

$$F + \frac{\partial F}{\partial t} = su + \beta\,\frac{\partial u}{\partial t} \quad \text{or}$$

$$F(t) = s_0 u(t) - \int_0^\infty u(t-\tau)\varphi(\tau)\,d\tau. \tag{6}$$

In these expressions α and β are material constants and $\varphi(\tau)$ is the so-called relaxation function. Usually $\varphi(\tau)$ consists of a sum of terms of the type $(D/\tau_R)\exp(-\tau/\tau_R)$, where D is a constant and τ_R the relaxation time, which may be anywhere between nanoseconds and days. The second version of Eq. (6) goes back to Boltzmann. As well as the first expression it represents a material that has memory; that is, the instantaneous value of $F(t)$ depends not only on the instantaneous value of the displacement $u(t)$ but also on the previous history $u(t-\tau)$.

Other loss mechanisms are dry (Coulomb) friction in the interface between two structural elements and radiation damping, which is caused by waves that propagate into the surrounding medium (e.g., air, water, sand, ground, etc.). See also Sections 3.5 and 4.3 in Chapter 11.

Equations (6) lead to lengthy expressions when they are used to derive equations of motion, but they still give linear expressions because no products or powers of the field quantities appear.

Mechanical vibrations are very often expressed as spectra; that is, the time dependence is written as a sum of terms of the type

$$\xi(t) = \mathrm{Re}\{\underline{A}_n c^{j\omega_0 nt}\}. \tag{7}$$

Here $\underline{A}_n$ is the complex amplitude (spectral value) and ω_0 the fundamental frequency. With this type of representation it is very convenient to introduce the concept of complex moduli, that is,

$$\underline{s} = s(1+j\eta), \qquad \underline{E} = E(1+j\eta), \qquad \underline{G} = G(1+j\eta), \tag{8}$$

where the underbars indicate that the quantity is complex ($j = \sqrt{-1}$). The term η is the loss factor. For many applications it can be assumed to be independent of frequency.

The great advantage of the loss factor is that loss mechanisms can be taken into account simply by replacing the real-valued elastic moduli by complex ones. It should be kept in mind, however, that complex moduli are an artefact, which works only for harmonic motions or sums of harmonic motions. The transformation of frequency spectra to the time domain that may be desirable in shock applications is problematic. The loss factor can also be expressed in terms of energies, as shown in Section 1 in Chapter 63. For more on the loss factor and its relation to other quantities, see Chapter 73 and Table 3 in Chapter 72.

2.3 Range of Validity

The relations given in Sections 2.1 and 2.2 can be used to find the equations of motion of mechanical structures, and if the appropriate boundary and initial conditions are used, they can be applied to practical situations. But even though many solutions of interest can be found this way, there remain also open questions. Usually they have to do with nonlinear effects or with time-varying parameters (e.g., stiffness). The causes for nonlinearities are (see also Chapter 65) as follows:

- Material nonlinearities; that is, the quantities $s, E, G, b, \alpha, \ldots$ in Eqs. (5) and (6) depend on amplitude, as may happen for very large vibrations or for strong shocks.
- Frictional forces that increase very strongly with amplitude; this effect is very often purposely applied in shock absorbers.
- Geometric nonlinearities, which occur when there is no longer a linear relation between the length change of a spring and the displacements of the adjacent bodies. Hertzian contact,[4] where the contact area changes with the displacement, and large-amplitude vibrations of thin plates,[5] where the length variations of the neutral plane can no longer be neglected, are examples of this type.
- Constraints at boundaries that cannot be expressed by linear equations; an example of this type is dry Coulomb friction, or motion between free clearances.

- Convective forces that are neglected in the inertia terms of Eqs. (1) and (2).

The consequence of slight nonlinearities is the existence of higher harmonics that are not contained in the excitation. When the nonlinearity is very pronounced, chaotic behavior may occur (see also Chapter 65). Examples for parameter changes (see also Section 3 in Chapter 65) are as follows:

- Length change of a pendulum
- Stiffness change due to rapid variations of the point of contact (e.g., teeth in a gearwheel, which can be considered as short cantilever beams that are loaded at different points)
- Impedance variations "seen" by a body that moves with constant speed over periodic supports (e.g., wheel on rail)

3 ENERGY RELATIONS AND VARIATIONAL PRINCIPLES

3.1 Hamilton's Principle and Rayleigh's Principle

In the previous section the motion of mechanical systems was expressed in terms of forces and displacements or similar dynamical and kinetic quantities. An alternative approach is to describe such systems by their energies and their displacements. The most powerful method in this type of calculation is Hamilton's principle,

$$\delta \int_{t_1}^{t_2} (W_{\text{kin}} - W_{\text{pot}})\, dt = 0, \qquad (9)$$

which states that "for an actual motion of the system, under the influence of the conservative forces, when it is started according to any reasonable initial conditions, the system will move so that the time average of the difference between kinetic and potential energies will be a minimum (or in a few cases, a maximum)" (Ref. 6, p. 281). In Eq. (9) W_{kin} and W_{pot} are the total kinetic and potential energies of a system. The symbol δ indicates that the variation has to be taken. For a few cases the potential and kinetic energies for surface or line elements are given in Table 1 in Chapter 64. If loss mechanisms have to be included, one can again employ the concept of the loss factor. As an alternative one can extend Eq. (9) by a dissipation function (see Ref. 7, Sections 81 and 82), which is a quadratic expression involving the relative velocities. Another extension of Eq. (9) can be made by introducing the work done by outside sources; this way all types of linear, forced vibrations can be treated.

Hamilton's principle is an extremely useful relation in vibroacoustics. It can be used as a starting point for the finite element method (FEM).[8,9] By employing the calculus of variations, it is also possible to derive equations of motion from Eq. (9),[10] to calculate the vibrations of coupled systems, and to find the phase speeds of different wave types.[11] (See also Refs. 12 and 13.)

A special solution of Eq. (9) is $\overline{W}_{\text{kin}} = \overline{W}_{\text{pot}}$, where the overbar denotes the time average. This relation is known as Rayleigh's principle[13] (see also Ref. 7, Section 88). It states that at resonance the time averages of kinetic and potential energies are equal. Since usually

$$\overline{W}_{\text{kin}} \sim \overline{\left(\frac{\partial \xi_i}{\partial t}\right)^2} \sim \omega^2 \overline{\xi_i^2}, \qquad (10)$$

Rayleigh's principle is a very convenient way of finding resonances. Its main advantage, however, is that a first-order error in the assumed spatial distribution of the displacements ξ_i (which are needed to find W_{kin} and W_{pot}) leads only to a second-order error in the resonance frequencies, provided that the boundary conditions are fulfilled. It also is known that Rayleigh's principle gives always an upper bound for the resonance frequency. Finding approximations for the lowest resonance frequency is just one application of Rayleigh's principle. It also can be used to find higher resonances and mode shapes (Rayleigh–Ritz method).[12,13]

As an example one may make the (wrong) assumption that the displacement of a simply supported beam in bending motion is given by $\xi_2 = 1 - 4x_1^2/l^2$. If one then uses the formulas in the appendix to Chapter 64 for a Bernoulli–Euler beam, Rayleigh's principle yields an approximate resonance frequency that is only a few percent higher than the correct value.

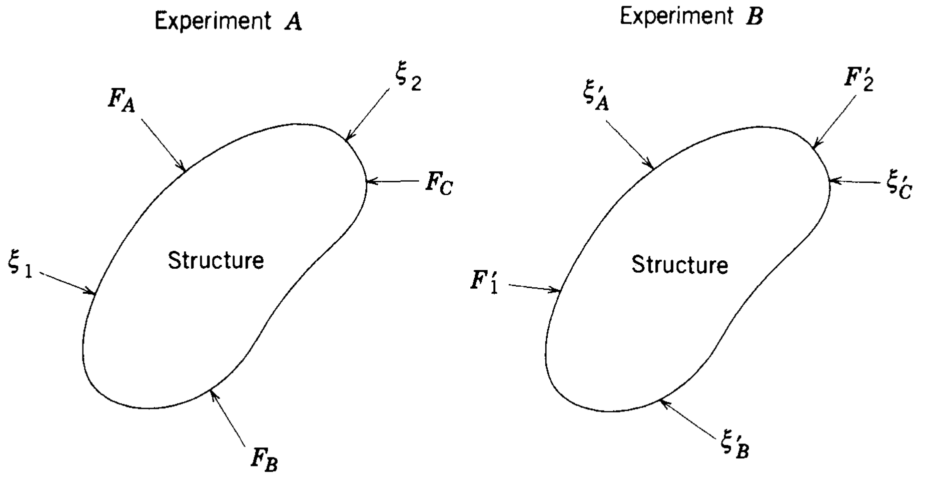
For the application of Rayleigh's principle to find the first stopping and passing bands of a periodic system, the reader is referred to Ref. 14.

3.2 Lagrange's Equation

Applying the calculus of variations to Hamilton's principle yields Lagrange's equation of the second type:

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{\xi}_n} - \frac{\partial L}{\partial \xi_n} = 0, \qquad L = W_{\text{kin}} - W_{\text{pot}}. \qquad (11)$$

Here ξ_n is the nth displacement coordinate and $\dot{\xi}_n$ the nth velocity coordinate. If forces F_n are driving the system from outside, the zero in the above equation has to be



$$F_A \xi'_A + F_B \xi'_B + F_C \xi'_C = F'_1 \xi_1 + F'_2 \xi_2$$

Fig. 2 Law of conjugate energies.

replaced by F_n. The main application of Eq. (11) is in the field of vibrations of multi-degree-of-freedom systems (see Chapter 63) where it can be used to calculate resonance frequencies, the shapes of eigenvectors, impulse responses, frequency responses, and so on.

3.3 Reciprocity

If the motion of the different parts of a mechanical system consists of small deflections about the equilibrium points, kinetic and potential energies are quadratic forms of these deflections. As a consequence, Hamilton's principle leads to symmetric differential equations and Eq. (11) is a set of symmetric, linear differential equations.* For such so-called self-adjoint expressions it is known that the principle of reciprocity holds (see, e.g., Ref. 15). In its simplest form this principle relates a force F_A acting at position x_A and generating a displacement ξ_1 at a point x_1 to a force F'_1 acting at x_1 and generating a displacement ξ'_A at x_A. It states

$$F_A \xi'_A = F'_1 \xi_1 \quad \text{or} \quad \frac{F_A}{\xi_1} = \frac{F'_1}{\xi'_A}. \tag{12}$$

Instead of ξ_1, ξ'_A one might also have used the velocities or accelerations. Care has to be taken that the directions of F_A and ξ_1 as well as F'_1 and ξ'_A are equal. The principle of reciprocity holds also for other pairs of field quantities, provided their product gives a power or energy, for example, angular velocity and moment (torque). In Section 6.3 in Chapter 11, the principle of reciprocity is used to relate the sound radiation from vibrating surfaces to the excitation of the same structure by a sound field.

An extension of the principle of reciprocity is Heaviside's law of conjugate energies[16] that can already be found in Ref. 7, Section 72. It is shown in Fig. 2. There F_A, F_B, ... are the forces acting at positions x_A, x_B, ... in the first experiment; ξ_1, ξ_2, ... are the displacements generated by them at the positions $x_1, x_2, \ldots$; $F'_1, F'_2, \ldots$ are the forces acting at x_1, x_2, ... in the second experiment; and ξ'_A, ξ'_B, ... are the displacements generated by them at x_A, x_B, ... The above equations hold for the complex amplitudes when the $\exp(j\omega t)$ notation is used.

Reciprocity is a very useful property. It can be used to calibrate transducers[17] and to identify unknown forces by making some auxiliary measurements.[18-20] For example, when F_A is unknown and unaccessible, one can measure ξ_1 and then make a measurement with an auxiliary force F'_1. When conjugate energies are used, more auxiliary measurements are necessary to get a set of linear equations for the unknown quantities.

4 GENERAL RESULTS

4.1 Frequency Spectra

The behavior of vibrating mechanical structures is quite often characterized by frequency spectra that determine the response for purely harmonic excitation of the type $\exp(j\omega t)$. Usually such spectra are given as the ratio of two (complex) amplitudes.

---

*In the literature one can also find nonsymmetric linear differential operators. They are the consequence of inconsistent approximations that are made in the derivation. Usually they give results that are very close to the correct ones, but the apparent violation of the principle of reciprocity is a certain drawback.

They are either measured by slowly sweeping through the frequency range of interest or are obtained as a result of a two-channel fast Fourier transform (FFT) measurement (transfer function). Typical examples are as follows:

$$\begin{array}{ll}
\text{Input mobility (admittance)} & \underline{A}_F = \mathbf{y}(x_0)/\mathbf{F}(\mathbf{x}_0) \\
\text{Input impedance} & \underline{Z}_F = \mathbf{F}(x_0)/\mathbf{v}(x_0) \\
\text{Apparent mass} & \mathbf{F}(x_0)/\mathbf{a}(x_0) \\
\textit{Dynamic stiffness} & \mathbf{F}(x_0)/\boldsymbol{\xi}(x_0) \\
\text{Transfer mobility} & \underline{A}_T = \mathbf{v}(x_1)/\mathbf{F}(x_0) \\
\text{Velocity ratio} & \mathbf{v}(x_1)/\mathbf{v}(x_2) \\
\text{Moment mobility} & \underline{A}_M = \mathbf{w}(x_0)/\mathbf{M}(x_0) \\
\text{Reflectioncoefficient} & \mathbf{v}_r/\mathbf{v}_j
\end{array} \qquad (13)$$

Obviously many other combinations are possible.

Some of the above quantities are of great importance because they can be related to the energy flow, as is shown by Eq. (25). Examples are

$$P = \frac{1}{2}\,|\mathbf{F}(x_0)|^2\,\mathrm{Re}\{\mathbf{A}_F\},$$

$$P = \frac{1}{2}\,|\mathbf{v}(x_0)|^2\,\mathrm{Re}\{\mathbf{Z}_F\},$$

$$P = \frac{1}{2}\,|\mathbf{M}(x_0)|^2\,\mathrm{Re}\{\mathbf{A}_M\},$$

$$\alpha = 1 - \left\|\frac{\mathbf{v}_r}{\mathbf{v}_i}\right\|^2. \qquad (14)$$

Here, $\mathbf{F}(x_0)$, $\mathbf{M}(x_0)$ are the driving force or moment at the excitation point; $\mathbf{v}$, $\mathbf{a}$, $\boldsymbol{\xi}$ represent velocity, acceleration, and displacement either at the excitation point x_0 or at some other point x_1 or x_2, $\mathbf{w}(x_0)$ is the angular velocity at the excitation point, $\mathbf{v}_i$, $\mathbf{v}_r$ are the amplitudes of the waves that are incoming to or reflected from a discontinuity; P is the mechanical power that is transmitted into a structure; and α is the absorption coefficient. To emphasise their complex nature, the quantities are underlined. Examples of, for example, spectra and mobilities can be found in the following chapters.

When the spectral behavior of a mechanical system is known over the whole range of interest, the above quantities can be used to find time histories. As an example, we consider a force $\mathbf{F}_{x0}(t)$ with period T and fundamental frequency $\omega_0 = 2\pi/T$. If this force is acting at x_0 and if the spectrum of the transfer mobility $\mathbf{A}_T$ is known, the time dependence of the velocity $v_{x1}(t)$ at point x_1 is given by

$$v_{x1}(t) = \frac{1}{2\pi} \sum_n \mathbf{F}^v_{x0}(n\omega_0) c^{jn\omega_0 t}/\mathbf{A}_T(n\omega_0) \qquad (15)$$

with

$$\mathbf{F}^v_{x0}(n\omega_0) = \frac{1}{T} \int_0^T \mathbf{F}_{x0}(t) c^{-jn\omega_0 t}\, dt.$$

The extension to nonperiodic signals follows the same lines as in other areas of acoustics; see Chapters 99–101.

4.2 Green's Function: Impulse Response

The basic idea of the application of Fourier analysis is to express all quantities as a sum of purely harmonic signals, or when a space dependence is also involved as a sum of plane; monochromatic, progressive waves. Thus these simple motions are used as "building blocks" to generate more complicated solutions. The great advantage of this approach is that, because the response of a system to purely harmonic time dependence or to plane-wave excitation can quite often be found, it takes only a summation to get the general solution (for the plane-wave case see wave impedances or transmission impedances and wave number spectra in Section 2.5 in Chapter 64).

An alternative to this "Fourier approach" is the use of Green's functions. In this case the response of a system to an idealized pulse is used as a building block for more complicated problems.

A Green's function $g(x,t|x_0,t_0)$ or impulse response function is defined by

$$L\{g(x,t|x_0,t_0)\} = A\delta(x - x_0)\delta(t - t_0). \qquad (16)$$

Here $L\{\cdot\}$ is the linear differential equation (or system of such equations) that describes the mechanical structure that is tested; $\delta(\cdot)$ is Dirac's delta function; and A is a constant that usually has a dimension (e.g., momentum transmitted by the pulse). Equation (16) shows that the Green's function gives the response (velocity, displacement, etc.) of a structure at point x and time t when it is excited by a pulse at point x_0 and time t_0.

Equation (16) is the general definition of Green's functions. A special case is when excitation point x_0 and observation point x are identical. In this case x and x_0 do not appear explicitly. For example, in other problems one is looking only for the response of structures when they are excited by a purely harmonic point force, etc.

In this case all quantities are proportional to $\exp(j\omega t)$ and the time t does not appear explicitly. For examples, see Eqs. (27a)–(27d) in Chapter 64. The Green's function approach has great advantages because any excitation function [e.g., the outside pressure $p(x,t)$ or a shock] can be expressed as

$$p(x,t) = \int p(x_0,t_0)\delta(x - x_0)\delta(t - t_0)\,dx_0\,dt_0. \tag{17}$$

In physical terms this means that any pressure distribution can be written as a sum of short, localized pulses of magnitude $p(x_0,t_0)\,dx_0\,dt_0$. Because of linearity, one can conclude that the response to any excitation distribution consists of the sum of the responses (Green's functions) of the short localized pulses; that is,

$$v(x,t) = \frac{1}{A}\int p(x_0,t_0)g(x,t|x_0,t_0)\,dx_0\,dt_0. \tag{18}$$

In this equation the response is expressed as a velocity, but it might also be another field quantity provided the proper Green's function is used. The constant A, which in the literature very often is set equal to unity, is retained here, because it makes it easier to keep track of the dimensions. The operation defined by Eq. (18) is also known as convolution, or *Faltung*.

Impulse response functions are very helpful when the response to transient excitation (e.g., shocks) is of interest. Shock design to a large degree consists in finding useful approximations for impulse response functions of complicated systems. See also Chapter 67.

Since Green's functions (or sets of Green's functions for different x and x_0) are solutions of the equation of motion of a structure, they contain all the information that is relevant for the vibration of the structure. That makes them very suitable for numerical problems when time histories of a motion have to be calculated. Similarly, they are very useful for problems where a linear structure is coupled to a localized nonlinear device (see e.g., Ref. 21).

The Green's function approach and Fourier transforms are just two sides of the same coin. A strong connection between them exists because the Fourier transform of a transfer function for normalized excitation is equal to the Green's function and vice versa. This fact is often used to find Green's functions. See Eq. (20).

4.3 Normal Modes

It is known (see, e.g., Refs. 12, 13, and 22 for applications or 6 and 23 for the mathematical foundations) that the vibrations of lightly damped mechanical structures can be expressed as

$$v(x) = \sum_{n=1}^{\infty} \frac{A_n\varphi_n(x)}{\omega_n^2(1 + j\eta_n) - \omega^2}. \tag{19}$$

Here $v(x)$ is the velocity (it might also be another quantity), A_n are constants that depend on the excitation and on mechanical properties, ω_n are the eigenfrequencies, η_n the (modal) loss factors, and $\varphi_n(x)$ the eigenfunctions (modes). Equation (19) holds for sinusoidal excitation. By applying a Fourier transform, it can be shown that the modal approach can also be used to describe the motion after the excitation has stopped, for example, after a short pulse or a shock. The result is

$$v_1(x,t) = \sum_n \frac{I_n}{\omega_n}\varphi_n(x)e^{-\eta_n\omega_n t/2}\sin\omega_n t. \tag{20}$$

Here I_n depends on the location and the type of excitation. Equation (20) shows that the free motion consists of several ringing modes that decay according to their modal loss factors. When the system is excited by a very short pulse, the total vibration consists of many modes. For longer pulses and for most shock applications, the ringing is concentrated to a few (or only one) low-frequency modes.

The eigenfunctions $\varphi_n(x)$ are solutions of the equation of motion for vanishing excitation. They have to fulfil the boundary conditions for the problem to be investigated. Equation (19) holds if the internal damping is small and if the energy flow through the boundaries can be neglected. As mentioned before, ω_n are the resonance frequencies. They determine the ringing sound after a structure has been excited (e.g., by a pulse). The $\varphi_n(x)$ are the vibration patterns or mode shapes that belong to these resonances. For many purposes it is very helpful to know that the eigenfunctions have the orthogonality property:

$$\int m''(x)\varphi_n(x)\varphi_m(x)\,dx = \begin{cases} 0 & \text{for } n \neq m, \\ A_n & \text{for } n = m. \end{cases} \tag{21}$$

Here the integration has to be taken over the whole structure. The term $m''(x)$ is the mass density. It is known[22] that at high frequencies the number ΔN of modes in a frequency band $\Delta\omega$ is given by

$$\frac{\Delta N}{\Delta\omega} = \frac{2m}{\pi}\,\mathrm{Re}\left\{\frac{1}{Z_F}\right\}. \tag{22}$$

Here m is the total mass of the system and $\mathbf{Z}_F$ its input impedance for point force excitation when there would be no reflection from boundaries. Thus $\mathbf{Z}_F$ is the impedance of the corresponding infinite system (see also Section 2.7 in Chapter 64). Apart from a few exceptions (see Section 2.9 in Chapter 64), eigenfunctions and eigenfrequencies cannot be represented by simple analytical functions. For most applications they have to be found numerically, usually by the FEM. (See, e.g., Refs. 8 and 9.) The great advantage is that once the φ_n and ω_n are known, the vibrations of a structure can in general be expressed in a very concise way by Eq. (19) because for many applications a few dozen modes (or less) give a rather accurate description (although thousands of elements may have been used to calculate them originally). Another advantage of the modal approach comes up in the high-frequency range where Eq. (21) is very helpful to obtain average values. This topic is treated more thoroughly in connection with statistical energy analysis (SEA) in Chapter 78.

TABLE 1 Material Properties[a]

Material	ρ (kg/m^3)	E ($\times 10^9$ N/m^2)	G ($\times 10^9$ N/m^2)	μ	c_L (m/s)	c_T (m/s)
Aluminum	2,700	72	27	0.34	5,160	3,160
Brass	8,400	100	38	0.36	3,450	2,100
Copper	8,900	125	48	0.35	3,750	2,300
Gold	19,300	80	28	0.43	2,030	1,200
Iron	7,800	210	80	0.31	5,180	3,210
Lead	11,300	16	5.6	0.44	1,190	700
Magnesium	1,740	45	17	0.33	5,080	3,100
Nickel	8,860	200	76	0.31	4,800	2,960
Silver	10,500	80	29	0.38	2,760	1,660
Steel	7,800	200	77	0.3	5,060	3,170
Tin	7,290	54	20	0.33	2,720	1,670
Zinc	7,140	100	41	0.25	3,700	2,400
Perspex	1,150	5.6	—	—	2,200	—
Polyamid	1,100–1,200	3.6–4.8	—	—	1,800–2,000	—
Polyethylene	≈900	3.3	—	—	1,900	540
Rubber[b]						
30 shore	1,010	0.0017	—	≈0.5	41	
50 shore	1,110	0.0023	—	≈0.5	46	
70 shore	1,250	0.0046	—	≈0.5	61	
Polivinylchloride	1,050	0.005–3.0	—	—	69–1,650	
Asphalt	1,800–2,300	7–21	—	—	1,900–3,100	
Brick						
Solid	1,800–2,000	≈16	—	—	2,600–3,100	
Hollow	700–1,000	3.1–7.8	—	—	2,100–2,800	
Concrete						
Dense	2,200–2,400	≈26	—	—	3,400	
Poured						
Concrete						
Light	1,300–1,600	≈4.0	—	—	1,700	
Porous	600–800	≈2.0	—	—	1,700	
Cork	120–250	0.02–0.05	—	—	430	
Fir	440–700	1.0–4.0	—	—	1,400–3,500	
Glass	2,500	≈4.4	—	—	4,200	
Gypsum board	700–950	4.1	—	—	2,000–2,500	
Oak	700–950	1.8–7.2	—	—	1,500–3,000	
Chipboard	600–750	≈4.9	—	—	≈2,700	
Plaster	≈1,700	≈4.3	—	—	≈1,600	
Plywood	600	5.4	—	—	3,000	

[a] For loss factors see Fig. 8, Chapter 71 of Volume 1.

[b] Dimension of test sample smaller than wavelength. Large rubber samples have c_L = 1500 m/s.

5 LIST OF MATERIAL PROPERTIES

For mechanical vibrations the most important material properties are density, elastic moduli, and the loss factors. In Table 1 some values of moduli are given. Since for many applications it is useful to know the wave speeds, the shear wave speed $c_T = \sqrt{G/\rho}$, and the longitudinal wave speed in rods, $c_L = \sqrt{E/\rho}$ is also given. Data on loss factors are shown in Fig. 8 in Chapter 71. Since loss factors vary with load, temperature, preconditioning, (sometimes) frequency-only ranges are given in the figure.

REFERENCES

1. L. Brekhovskikh and V. Goncharov, *Mechanics of Continua and Wave Dynamics*, Springer, Berlin, 1985, Chapters 1–4.
2. J. D. Achenbach, *Wave Propagation in Elastic Solids*, North-Holland, Amsterdam, 1973.
3. C. Zener, *Elasticity and Anelasticity of Metals*, University of Chicago Press, Chicago, 1948.
4. K. L. Johnson, *Contact Mechanics*, Cambridge University Press, Cambridge, 1985, Chapters 4 and 7.
5. S. P. Timoshenko and S. Woinowsky-Krieger, *Theory of Plates and Shells*, McGraw-Hill, New York, 1959, Chapter 13.
6. P. M. Morse and H. Feshbach, *Methods of Theoretical Physics*, Vol. 1, McGraw-Hill, New York, 1953, Chapter 3.2.
7. Lord J. W. Rayleigh, *The Theory of Sound*, Vol. I, Macmillan, London, 1894, or Dover, New York, 1945.
8. M. Petyt, "Finite Element Techniques for Structural Vibration," and "Finite Element Technique for Acoustics," in R. G. White and J. G. Walker (Eds.), *Noise and Vibration*, Wiley, New York, 1982, Chapters 15 and 16.
9. O. Zienkiewicz, *The Finite Element Method*, McGraw Hill, London, 1977.
10. M. C. Junger and D. Feit, Sound, Structures and Their Interaction, MIT Press, Cambridge, 1972, Chapter 9.
11. M. Heckl, "Einfache Anwendung des Hamilton'schen Prinzips bei Kö rperschallproblemen," *Acustica*, Vol. 72, 1990, pp. 189–196.
12. L. Meirowitch, *Elements of Vibration Analysis*, McGraw-Hill, New York, 1986.
13. R. E. D. Bishop and D. C. Johnson, *The Mechanics of Vibration*, Cambridge University Press, London, 1979, Chapters 3.9, 5.3, 7.10.
14. D. J. Mead, "Response of Periodic Structures to Noise Fields," in *Noise and Vibration*, R. G. White and J. G. Walker (Eds.), Wiley, New York, 1982, Chapter 13.
15. L. M. Lyamshev, "A Question in Connection with the Principle of Reciprocity in Acoustics," *Sov. Phys. Doklady*, Vol. 4, 1959, pp. 406–409.
16. O. Heaviside, *Electrical Papers*, Vol. I, p. 520 and Vol. II, p. 202.
17. L. L. Beranek, *Acoustic Measurements*, Wiley, New York, 1949, Chapter 4.
18. T. ten Wolde and G. Gadefelt, "Development of Standard Measurements for Structure-Borne Sound," *Noise Control Engr.*, Vol. 28, 1987, pp. 5–15.
19. M. Heckl, "Anwendungen des Satzes von der wechselseitigen Energie," *Acustica*, Vol. 58, 1985, pp. 111–117.
20. I. L. Vier, "Use of Reciprocity and Superposition in Predicting Power Input to Structures Excited by Complex Sources," *Proceedings INTER-NOISE*, Newport Beach, California, 1989, Vol. 89, pp. 543–546.
21. M. E. McIntyre, R. T. Schumacher, and J. Woodhouse, "On the Oscillations of Musical Instruments," *J. Acoust. Soc. Am.*, Vol. 74, 1983, pp. 1325–1345.
22. L. Cremer, M. Heckl, and E. E. Ungar, *Structure-Borne Sound*, 2nd ed., Springer, Berlin, 1988.
23. R. Courant and D. Hilbert, *Methoden der mathematischen Physik I*, Springer, Berlin, 1931, Chapters V, VI.

63

EXPERIMENTAL AND THEORETICAL STUDIES OF VIBRATING SYSTEMS

PAUL J. REMINGTON

1 INTRODUCTION

This chapter provides an overview and introduction to analytical and experimental methods applied to vibrating systems. The initial focus is on the single-degree-of-freedom oscillator, not only because it is simple and easy to understand but also because it is a useful idealized model of many complicated vibrating systems. The simple oscillator serves as a vehicle for the discussion of free and forced response, Fourier transform methods, impulse response functions, energy balance, and equivalent circuits. The focus then shifts to more complicated multi-degree-of-freedom lumped-parameter systems where the concepts of mode shape functions and natural frequencies are introduced. The chapter concludes with a discussion of experimental methods. Similar to the discussion of analytical methods, preceding it, the discussion of experimental methods begins with the single degree-of-freedom oscillator and its use in the determination of material properties. The chapter concludes with a discussion of the extension of these techniques to multi-degree-of-freedom systems, commonly referred to as experimental modal analysis.

2 SINGLE-DEGREE-OF-FREEDOM SYSTEMS

2.1 Free Vibration

Equations of Motion While very simple, the linear, single-degree-of-freedom oscillator depicted in Fig. 1 is a very useful idealized model of many real-world vibrating systems. A simple example is an engine on elastic mounts where the engine represents the mass and the elastic mounts the spring and damper. Another example is a piezoelectric accelerometer in which a small seismic mass is mounted on a slightly resilient piezoelectric crystal. The main reason, however, for treating single-degree-of-freedom oscillators rather extensively here is the fact that in the neighborhood of resonances many systems, especially when they are lightly damped, behave very much like single-degree-of-freedom systems even though they may have many resonances.

In Fig. 1 the displacement of the mass m is denoted by u. This motion is assumed to be in a straight line with no rotation. Thus the inertial force is

$$m\,\frac{d^2u}{dt^2}.$$

The force acting on the damper element, which is assumed to be viscous in nature, is given by

$$b\,\frac{du}{dt},$$

where b is the damping coefficient and the spring force is ku, where k is the spring constant. When the three relations are combined and an outside force $f(t)$ is added as shown in Fig. 1, the equation of motion

$$m\,\frac{d^2u}{dt^2}+b\,\frac{du}{dt}+ku=f(t) \qquad (1)$$

of a single-degree-of-freedom system is obtained.

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker

Copyright © 1997 John Wiley & Sons, Inc.

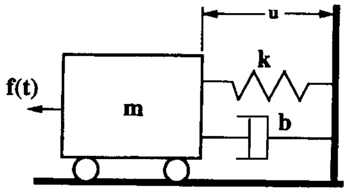


Fig. 1 Single-degree-of-freedom oscillator.

Response Two distinct solutions can be found to Eq. (1), the free response and the forced response. The free response (transient response) of the system is the response that occurs in the absence of any external force. In this section the free response will be considered by setting the external force $f(t)$ in Eq. (1) to zero and computing the resulting motion of the mass when it is given an initial displacement and velocity and then released. To keep things simple, the undamped case when $b = 0$ will be considered first. Although there is some damping in all real physical systems, the undamped case is a reasonable approximation to very lightly damped systems. If the mass is displaced a distance u_0 from its equilibrium position and is released with an initial velocity du_0/dt, then the motion of the mass as a function of time becomes

$$u = A_0 \cos(\omega_0 t + \phi_0), \tag{2}$$

where

$$A_0 = \left[\left(\frac{du_0/dt}{\omega_0}\right)^2 + u_0^2\right]^{1/2},$$

$$\phi_0 = \tan^{-1}\left(\frac{-du_0/dt}{\omega_0 u_0}\right),$$

$$\omega_0 = \left(\frac{k}{m}\right)^{1/2}.$$

As Eq. (2) shows, the motion of the mass is a pure sinusoid of period $2\pi/\omega_0$. The quantity ω_0 is called the natural frequency of the system and in the undamped case is given by the very simple formula shown above.

If a small amount of damping is now added to the system, the free response of the mass will change to

$$u = A_d e^{-\zeta\omega_0 t} \cos(\omega_d t + \phi_d), \tag{3}$$

where

$$A_d = \left[u_0^2 + \frac{(du_0/dt + \zeta\omega_0 u_0)^2}{\omega_d^2}\right]^{1/2},$$

$$\phi_d = \tan^{-1}\left[\frac{-(du_0/dt + \zeta\omega_0 u_0)}{u_0\omega_d}\right],$$

$$\omega_d = \omega_0(1 - \zeta^2)^{1/2},$$

$$\zeta = \frac{1}{2}\frac{b}{(mk)^{1/2}} < 1,$$

and u_0 and du_0/dt, as before, are the initial displacement and velocity, respectively. As the equation indicates, when the mass is displaced from its equilibrium position, it will oscillate at frequency ω_d (slightly lower than ω_0) and the oscillations will decay exponentially with time. The rate of decay depends on the damping ratio ζ and the undamped natural frequency ω_0. In many practical instances the shift in natural frequency due to damping is very small and the motion of the mass is a very slowly decaying, nearly pure harmonic oscillation at the undamped frequency ω_0. Figure 2 illustrates the resulting motion when the mass is displaced u_0 from its equilibrium position and released with no initial velocity when the damping ratio ζ is 0.01.

Equation (3) is correct as long as ζ, the damping ratio, is less than 1. If ζ exceeds 1, no oscillations will occur, and when the mass is displaced from its equilibrium position, it will simply decay back to its equilibrium position with no oscillation according to the equation

$$u = -\frac{du_0/dt + \omega_0 u_0 b_2}{2\omega_d'} e^{-b_1\omega_0 t} + \frac{du_0/dt + \omega_0 u_0 b_1}{2\omega_d'} e^{-b_2\omega_0 t}, \tag{4}$$

where

$$b_1 = (\zeta^2 - 1)^{1/2} + \zeta,$$

$$b_2 = \zeta - (\zeta^2 - 1)^{1/2},$$

$$\omega_d' = \omega_0(\zeta^2 - 1)^{1/2}.$$

Figure 3 illustrates the response of an "overdamped" system ($\zeta = 1.05$) when, as in the previous example, the mass is displaced u_0 from its equilibrium position and released with no initial velocity. As the figure shows, no oscillations of the mass occur in this case, and the mass

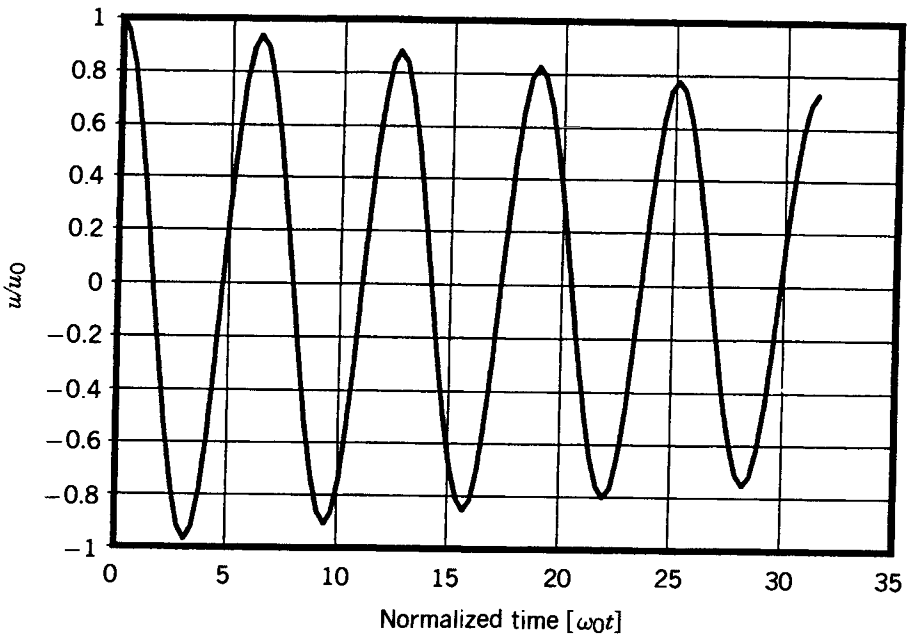


Fig. 2 Free response of a lightly damped single-degree-of-freedom oscillator.

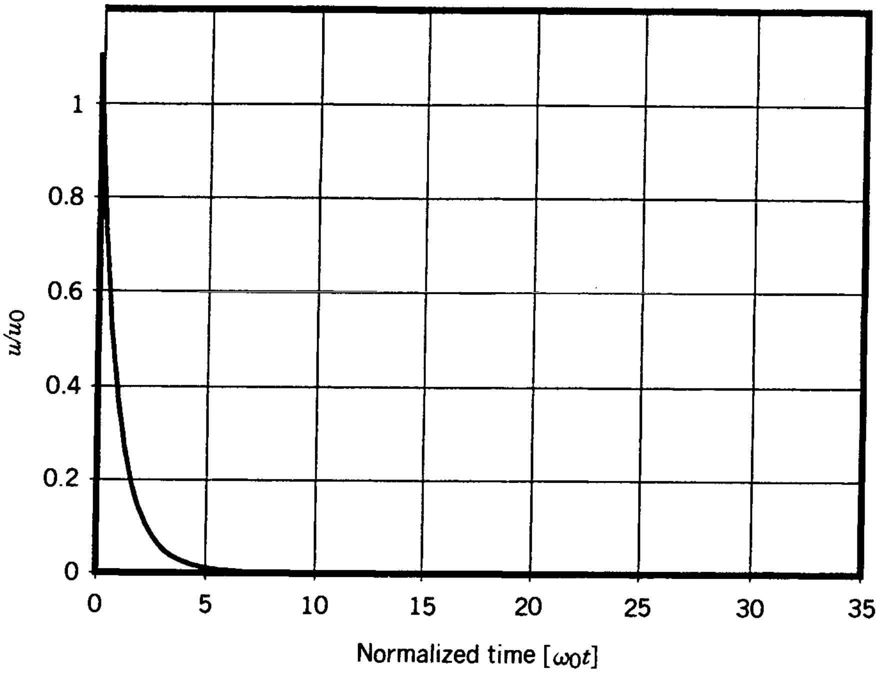


Fig. 3 Free response of an overdamped single-degree-of-freedom oscillator.

simply settles back to its equilibrium position. The *in-between* case in which $\zeta = 1$ or $b/2m = \omega_0$ is called critical damping. For such a system the motion of a mass initially displaced by u_0 and set into motion with velocity du_0/dt is given by

$$u = \left[u_0 + \left(\frac{du_0}{dt} + \omega_0 u_0 \right) t \right] e^{-\omega_0 t}. \tag{5}$$

Systems with critical damping and those that are overdamped do not occur very often in real life. However, in the field of servomechanisms and electromechanical positioning systems, they are of importance.

Energy The total energy associated with the oscillator is given by the sum of the kinetic energy in the mass plus the potential energy in the spring:

$$W = \tfrac{1}{2}m\left(\frac{du}{dt}\right)^2 + \tfrac{1}{2}ku^2, \qquad (6)$$

which for the undamped system yields

$$W = \tfrac{1}{2}m\omega_0^2A^2\sin^2(\omega_0t+\phi_0) + \tfrac{1}{2}kA^2\cos^2(\omega_0t+\phi_0),$$

or on simplifying the equation with $\omega_0^2 = k/m$,

$$W = \tfrac{1}{2}kA^2 = \tfrac{1}{2}m(\omega_0A)^2. \qquad (7)$$

As the equation shows the total energy in the system is equal to the kinetic energy when the mass attains its greatest speed or to the potential energy when the spring attains its greatest deflection.

If we multiply Eq. (1) by du/dt, we find that

$$m\frac{d^2u}{dt^2}\frac{du}{dt} + ku\frac{du}{dt} + b\left(\frac{du}{dt}\right)^2 = f(t)\frac{du}{dt},$$

which is identical to

$$\frac{d}{dt}\left[\frac{1}{2}m\left(\frac{du}{dt}\right)^2 + \frac{1}{2}ku^2\right] + b\left(\frac{du}{dt}\right)^2 = f(t)\frac{du}{dt}.$$

Recognizing the quantity in brackets as W, we can write

$$\frac{dW}{dt} + P_d = P_i, \qquad (8)$$

where P_d is the dissipated power and P_i is the power injected by the external force. For a lightly damped system we can use Eq. (7) to approximate the average energy $\overline{W}$ as

$$\overline{W} = \tfrac{1}{2}m\left(\frac{du}{dt}\right)^2_{\max} = m\overline{(du/dt)^2}, \qquad (9)$$

where the overbar means the time average over one cycle. The average dissipated power can be written similarly as

$$\overline{P}_d = b\overline{(du/dt)^2} = \frac{b}{m}\overline{W}. \qquad (10)$$

In the steady state where $dW/dt = 0$, Eq. (8) gives a simple energy balance that states that $P_d = P_i$ or that the dissipated power equals the injected power. As a result the time average dissipated power and injected power are related in the steady state by

$$\overline{P}_d = b\overline{(du/dt)^2} = \frac{b}{m}\overline{W} = \overline{P}_i. \qquad (11a)$$

In the transient state when $\overline{P}_i = 0$, the power balance in Eq. (8) gives

$$\frac{d\overline{W}}{dt} + \frac{b}{m}\overline{W} = \overline{P}_i \quad \text{or} \quad \overline{W} = W_0e^{-(b/m)t} = W_0e^{-2\zeta\omega_0t}. \qquad (11b)$$

In this equation $\overline{W}$ is a slowly varying function because it has been averaged over one (or more) cycles.

Instead of the damping coefficient b or the damping ratio ζ, the loss factor η is often used. It is defined by

$$\eta = \frac{1}{2\pi}\left(\frac{W_{\text{cycle}}}{W_{\text{rem}}}\right), \qquad (12)$$

where W_{cycle} is the energy lost per cycle and W_{rem} is the energy remaining after the cycle. For lightly damped systems $W_{\text{rem}} \approx \overline{W}$ and $W_{\text{cycle}} = \overline{P}_dT$, where T is the time duration of one cycle of the system or $T = 2\pi/\omega_0$. Using these quantities in Eq. (12) we obtain the definition of the loss factor in terms of the damping ratio

$$\eta = \frac{b}{m\omega_0} = \frac{b}{(mk)^{1/2}} = 2\zeta. \qquad (13)$$

More information about damping can be found in Chapter 71.

Example As a simple example consider the point mass suspended at the end of a cantilever beam as shown in Fig. 4. It is assumed that the mass of the beam is much less than the suspended mass and that the rotary inertia of the suspended mass is too small to affect the motion of the beam. The force deflection relationship for the cantilever beam is given by

$$f = \left(\frac{3EI}{L^3}\right)u \quad \text{or} \quad k = \frac{3EI}{L^3},$$

where EI is the bending stiffness of the beam and L is the length, which gives, for the natural frequency of the mass m on the cantilever,

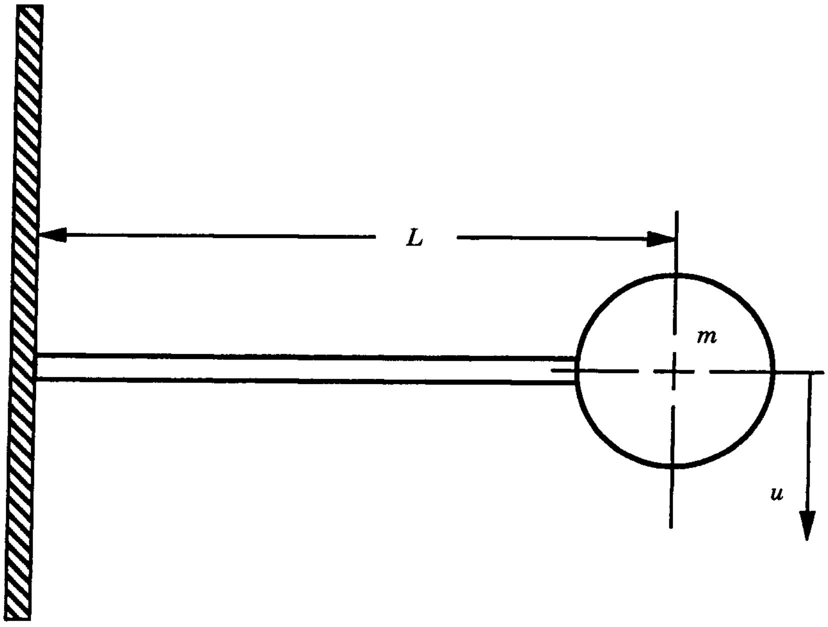


Fig. 4 Mass on the end of a cantilever beam.

$$\omega_0 = \left(\frac{k}{m}\right)^{1/2} = \left(\frac{3EI}{mL^3}\right)^{1/2} .$$

For a lightly damped system if the mass is displaced a distance u_0 from its equilibrium position and if it is released with no initial velocity, the position of the end of the beam as a function of time will be [see Eq. (3)]

$$\begin{aligned} u &= \frac{u_0}{(1-\zeta^2)^{1/2}}\, e^{-\zeta\omega_0 t} \cos(\omega_0\sqrt{1-\zeta^2}\,t + \phi_d) \\ &\approx u_0 e^{-\zeta\omega_0 t} \cos \omega_0 t, \qquad \zeta < 1, \\ \phi_d &= \tan^{-1}\left(\frac{-\zeta}{(1-\zeta^2)^{1/2}}\right) . \end{aligned} \tag{14}$$

As the equation shows, the free response of the beam will depend on the magnitude of the damping ratio ζ. Unfortunately, the analytical determination of damping in dynamic systems is usually very difficult. Consequently, it is usually estimated based on measurements. In fact, as will be described below, an experiment involving the natural decay of an oscillator, the same as that described by Eq. (14), can be used to estimate the damping.

2.2 Forced Response

Harmonic Forcing If the external force in Eq. (1) is not zero, but is suddenly applied to the mass, then the response will be a linear combination of the free or transient response, just described, and the forced response. In real dynamic systems the transient response quickly dies out due to damping, and one is left with only the forced response.[1] If a harmonic force of the form $\mathrm{Re}\{F_0 e^{j\omega_f t}\}$ (where $\mathrm{Re}\{\cdot\}$ means the real part of the quantity in brackets) is applied to the mass of Fig. 1, the forced displacement of the mass relative to its equilibrium point will be of the form

$$u(t) = \mathrm{Re}\{U e^{j\omega_f t}\},$$

where ω_f is the forcing frequency. In what follows we will use phasor notation and omit the symbol $\mathrm{Re}\{\cdot\}$.

The ratio of the complex displacement amplitude to the force amplitude can be calculated as[1,2]

$$\frac{U}{F_0} = \left\{k\left[1 - \left(\frac{\omega_f}{\omega_0}\right)^2 + \frac{2j\zeta\omega_f}{\omega_0}\right]\right\}^{-1} . \tag{15}$$

The ratio of system response amplitude to harmonic force amplitude as a function of frequency is often referred to as the frequency response function of the dynamic system. Both velocity and acceleration amplitude are commonly used in place of the displacement. A variety of different terminologies have been used over the years to name these various ratios and their inverses.

TABLE 1 Terminology for Frequency Response Functions

Response Measurement	Response/Force	Force/Response
Displacement	Receptance Compliance	Dynamic stiffness
Velocity	Admittance Mobility	Impedance
Acceleration	Accelerance Inertance	Apparent mass

Table 1 lists the ratios and some of the names associated with each. The reactance given in Eq. (15) will be used in sections to follow as will the admittance or mobility, which is given by

$$Y(\omega_f) = \frac{j\omega U}{F_0} = \left\{ \frac{k}{j\omega_f} \left[1 - \left(\frac{\omega_f}{\omega_0} \right)^2 + 2j\zeta \frac{\omega_f}{\omega_0} \right] \right\}^{-1},$$

or in terms of amplitude and phase

$$Y(\omega) = |Y(\omega)| e^{j\phi},$$

$$|Y(\omega_f)| = \left\{ \frac{k}{\omega_f} \right\}^{-1}$$

$$\cdot \left\{ \left[1 - \left(\frac{\omega_f}{\omega_0} \right)^2 \right]^2 + \left[2\zeta \left(\frac{\omega_f}{\omega_0} \right) \right]^2 \right\}^{-1/2},$$

$$\phi = \tan^{-1} \frac{1 - (\omega_f/\omega_0)^2}{2\zeta(\omega_f/\omega_0)}. \qquad (16)$$

Figure 5 illustrates the magnitude and phase of the admittance as a function of the forcing frequency ω_f. Figure 5 shows a peak in the admittance that occurs for very light damping at $\omega_f = \omega_0$, the natural frequency of the oscillator. The bandwidth of the resonance is also shown in Figure 5. It is measured at the point where the admittance drops by 3 dB and for a lightly damped oscillator is given approximately by

$$\delta\omega = 2\zeta\omega_0 = \eta\omega_0.$$

At the undamped natural frequency of a lightly damped oscillator, the admittance is given by

$$Y(\omega_f) = \{2\zeta\omega_0 m\}^{-1}, \qquad \omega_f = \omega_0.$$

The admittance is real and controlled by the magnitude of the damping. For forcing frequencies well below ω_0 the admittance is given approximately by

$$Y(\omega_f) \approx \frac{j\omega_f}{k}, \qquad \omega_f \ll \omega_0.$$

The admittance is now complex with a phase shift between velocity and force of 90° and appears to be approximately a pure stiffness. Finally for frequencies well above the natural frequency the admittance becomes

$$Y(\omega_f) \sim \{j\omega_f m\}^{-1}, \qquad \omega_f \gg \omega_0,$$

is once again purely imaginary with a phase shift now of −90°, and approximates a pure mass.

In the above the frequency response functions have all been given as if the damping were viscous, that is, assuming that the damping coefficient b is constant with frequency. In many practical cases the damping is better modeled as $b = c/\omega$, where c is a constant independent of frequency. This form of damping is referred to as *hysteretic* or *structural* damping. It is equivalent to assuming that the stiffness is not real but is complex and of the form $k = k_0(1 + j\eta)$.

For structural damping the admittance takes the form

$$Y(\omega_f) = \left\{ \frac{k}{j\omega_f} \left[1 - \left(\frac{\omega_f}{\omega_0} \right)^2 + j\eta \right] \right\}^{-1}.$$

Here η is the loss factor defined in Eq. (12) and described more fully in Chapter 71.

Equivalent Circuit It is often useful when visualizing complex arrangements of springs, masses, and dampers to utilize an equivalent electrical circuit representation of the system analogous to the equivalent electrical circuit discussed in Chapter 13 for acoustical systems. If the analogy is made that force is voltage and velocity is current, one obtains the representation shown in Fig. 6 for the oscillator of Fig. 1. The external force becomes a voltage source, the mass is equivalent to an inductor, the spring is analogous to a capacitor, and the damper is a resistor. While this approach offers little additional insight for the simple oscillator, in truly complex systems the equivalent circuit can be of significant help in visualizing the interactions of the various elements. In addition the equivalent circuit approach is not limited to lumped-parameter systems but can be extended to visualize the interactions between continuous systems (beams, plates,

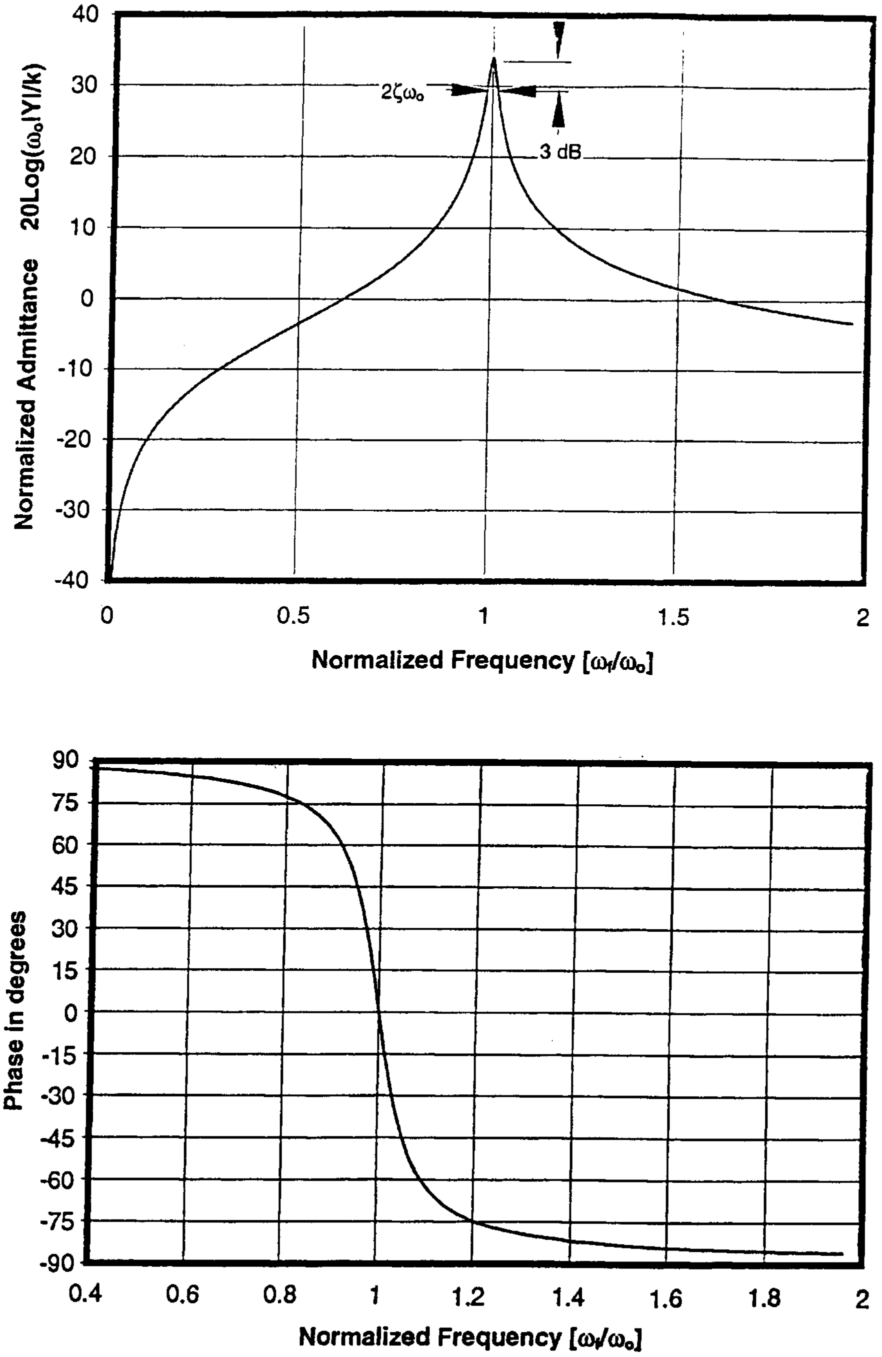


Fig. 5 Admittance of a single-degree-of-freedom oscillator with a damping ratio of 0.01.

acoustic volumes, etc.) by introducing frequency-dependent impedances. Figure 7 illustrates the equivalent circuit for a slightly more complex dynamic system. The simple oscillator is now attached to a finite beam, and the influence of the beam is included through the use of the point impedance of the beam at the point of attachment of the oscillator. It is interesting to note that the series arrangement of the spring–damper with the beam in the

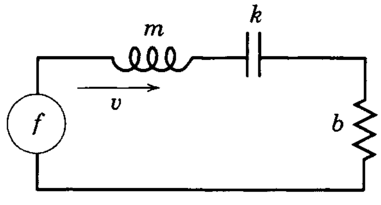


Fig. 6 Equivalent circuit for a spring–mass system.

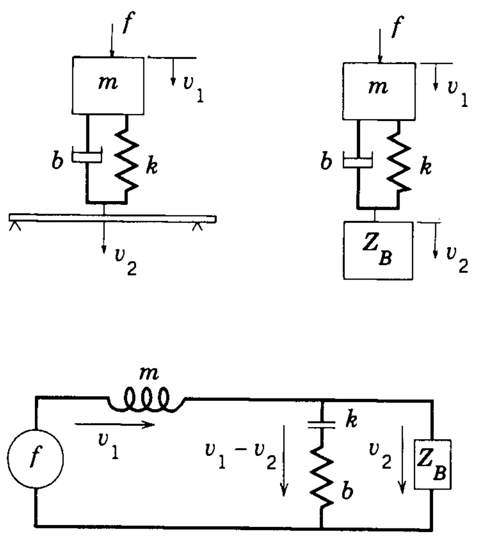


Fig. 7 Equivalent circuit example for a single-degree-of-freedom oscillator on a beam.

mechanical model results in a parallel arrangement in the equivalent circuit, and the parallel arrangement of spring and damper in the mechanical model results in a series arrangement in the equivalent circuit. In this equivalent circuit analogy, force is analogous to voltage and velocity is analogous to current. Under some circumstances it is more appropriate to formulate an equivalent circuit in which force is analogous to current and velocity is analogous to voltage. For that equivalent circuit analogy parallel or series arrangements of elements in the mechanical system will result in the same arrangement in the electrical system. (See Chapter 13.)

Fourier Transform Since the systems considered here are linear, one can apply the principle of superposition to use the frequency response derived above to compute the response of the system to forcing functions that are not harmonic. To do so one utilizes the Fourier transform and its inverse,[3] which are defined as

$$x(t) = \int_{-\infty}^{\infty} X(\omega)e^{j\omega t}\, d\omega, \tag{17}$$

$$X(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} x(t)e^{-j\omega t}\, dt, \tag{18}$$

where $x(t)$ is a real function and $X(\omega)$ is, in general, a complex function of ω. If one substitutes the Fourier transform representation of the external force $f(t) = \int_{-\infty}^{\infty} F(\omega)e^{j\omega t}\, d\omega$ and the Fourier transform of the velocity of the oscillator from its equilibrium position $du/dt = v(t) = \int_{-\infty}^{\infty} V(\omega)e^{j\omega t}\, d\omega$ into Eq. (1) and solves for $V(\omega)$ in terms of $F(\omega)$, one obtains

$$\begin{aligned} V(\omega) &= \left\{ j\omega m\left[1 - \left(\frac{\omega_0}{\omega}\right)^2 - 2j\zeta\left(\frac{\omega_0}{\omega}\right)\right]\right\}^{-1} F(\omega) \\ &= Y(\omega)F(\omega). \end{aligned} \tag{19}$$

An expression for the time response of the mass can then be obtained by taking the inverse Fourier transform of Eq. (19) as follows:

$$v(t) = \int_{-\infty}^{\infty} Y(\omega)F(\omega)e^{j\omega t}\, d\omega. \tag{20}$$

In Eq. (20) the external force has been broken up into a continuous distribution of infinitesimal forces of complex amplitude $F(\omega)\, d\omega$ at each frequency ω and the resulting velocity response computed by integrating the response to a force at a single frequency over all frequency so as to sum the contributions from each force. The function $F(\omega)$ can be obtained by utilizing the inverse Fourier transform of $f(t)$,

$$F(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(t)e^{-j\omega t}\, dt. \tag{21}$$

As an example consider a force of the form

$$f(t) = \begin{cases} F_0 e^{\beta t}, & t < 0, \\ F_0 e^{-\beta t}, & t \geq 0. \end{cases}$$

Substituting this function into Eq. (21) and carrying out the integration, one obtains

$$F(\omega) = \frac{F_0\beta}{\pi(\beta^2 + \omega^2)}.$$

If one substitutes this equation into Eq. (20) and carries out the indicated integration using contour integration methods,[3] one obtains

$$v(t) = \begin{cases} \dfrac{F_0\beta e^{\beta t}}{m(\beta^2 + \omega_0^2)}, & t < 0, \\ \dfrac{F_0\beta}{m(\beta^2 + \omega_0^2)}(-e^{-\beta t} + 2e^{-\zeta\omega_0 t}\cos \omega_0 t), & t > 0. \end{cases}$$

The result, which is valid for light damping ($\zeta^2 \ll 1$), shows that for $t < 0$ the velocity increases exponentially with time as the force increases. At $t = 0$ there is a discontinuity in the increase of the force, which excites the transient response of the oscillator (the second term in the equation for $t > 0$). As expected, however, the velocity, displacement, and acceleration are continuous at $t = 0$. Another approach to the estimation of the response of dynamic systems to nonharmonic forces is the use of the Laplace transform, a variation on the Fourier transform, which is especially useful for excitation that is zero up to $t = 0$. The interested reader is referred to Ref. 3 for further information.

Impulse Response (Green's Function) The Fourier transform is a frequency-domain approach. It is also possible to estimate dynamic system response to nonharmonic excitation in the time domain by using the impulse response function. For the simple oscillator of Fig. 1 the impulse response can be obtained from Eq. (3) for the transient response. It is only necessary to estimate the initial conditions. It is straightforward to show that if a unit impulse (delta function) is applied to the mass in Fig. 1 at $t = t_0$, at $t = t_0^+$ the initial displacement will be zero, and the initial velocity will be

$$\left(\frac{du_0}{dt}\right)_{t=t_0^+} = v_0 = \frac{1}{m}.$$

This equation does not mean that velocity has the units of $1/m$, where m is mass. Since we used a unit impulse, the 1 appearing in the equation has the units of momentum or mv.

By substituting this velocity into Eq. (3), one can obtain the displacement response of a lightly damped oscillator to a unit impulse as follows:

$$h(t - t_0) = \begin{cases} \dfrac{1}{m\omega_d} e^{-\zeta\omega_0(t-t_0)} \sin[\omega_d(t - t_0)], & t > t_0, \\ 0, & t < t_0 \end{cases} \tag{22}$$

Causality requires that physically realizable systems not respond before the impulse is applied. Consequently the impulse response function is zero for $t < t_0$. If one now takes a force $f(t_0)$, divides the force up into a number of infinitesimal impulses of magnitude $f(t_0)\,dt_0$, one can obtain the displacement of the oscillator at any time $t > t_0$ by integrating the product of the impulse amplitude and impulse response over all t_0 as follows:

$$u(t) = \int_{-\infty}^{\infty} h(t - t_0) f(t_0)\, dt_0 = \int_{-\infty}^{t} h(t - t_0) f(t_0)\, dt_0, \tag{23}$$

where the second form is equivalent to the first because $h(t - t_0) = 0$ for $t_0 > t$. Equation (23) is referred to as the convolution integral and has an equivalent form

$$u(t) = \int_{-\infty}^{\infty} h(t_0) f(t - t_0)\, dt_0 = \int_{0}^{\infty} h(t_0) f(t - t_0)\, dt_0.$$

If one takes the Fourier transform of both sides of Eq. (23), one obtains

$$U(\omega) = \frac{V(\omega)}{j\omega} = H(\omega) F(\omega), \tag{24}$$

where

$$H(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} h(t) e^{j\omega t}\, dt$$

is the Fourier transform of the impulse response. Here, $U(\omega)$, $V(\omega)$, and $F(\omega)$ are the Fourier transforms of the displacement, velocity, and force, respectively. By comparing Eq. (24) and Eq. (19), it is clear that

$$H(\omega) = \frac{Y(\omega)}{j\omega}$$

or that the Fourier transform of the impulse response is the frequency response.

3 MULTI-DEGREE-OF-FREEDOM SYSTEMS

3.1 Basic Equations

Figure 8 shows a simplified multi-degree-of-freedom lumped-parameter oscillator. One equation of the form

$$m_i \frac{d^2 u_i}{dt^2} - b_i \frac{du_{i+1}}{dt} + (b_i + b_{i-1}) \frac{du_i}{dt} - b_i \frac{du_{i-1}}{dt} - k_i u_{i+1} + (k_i + k_{i-1}) u_i - k_i u_{i-1} = f_i \tag{25}$$

can be written for each mass in the system. The resulting N equations in the N unknown displacements u_i can be written in matrix form to yield

$$[M]\left\{\frac{d^2 u}{dt^2}\right\} + [B]\left\{\frac{du}{dt}\right\} + [K]\{u\} = \{f\}, \tag{26}$$

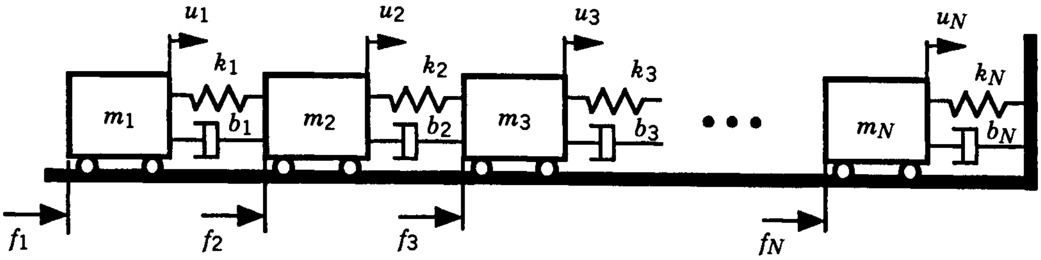


Fig. 8 Example of a multi-degree-of-freedom oscillator.

where the mass matrix is given by

$$[M] = \begin{bmatrix} m_1 & 0 & 0 & \cdots & 0 \\ 0 & m_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & m_N \end{bmatrix},$$

the damping matrix is given by

$$[B] = \begin{bmatrix} b_1 & -b_1 & 0 & 0 & \cdots \\ -b_1 & b_1 + b_2 & -b_2 & 0 & \cdots \\ 0 & -b_2 & b_2 + b_3 & -b_3 & \cdots \\ 0 & \ddots & & & \\ \vdots & & & & \end{bmatrix},$$

and the stiffness matrix is given by

$$[K] = \begin{bmatrix} k_1 & -k_1 & 0 & 0 & \cdots \\ -k_1 & k_1 + k_2 & -k_2 & 0 & \cdots \\ 0 & -k_2 & k_2 + k_3 & -k_3 & \cdots \\ 0 & \ddots & & & \\ \vdots & & & & \end{bmatrix}.$$

All matrices are square and symmetric. The notation $\{d^2u/dt^2\}$, $\{du/dt\}$, $\{u\}$, and $\{f\}$ means that these quantities are column vectors containing the acceleration, velocity, displacement, and force, respectively, associated with each mass. In general the three square matrices are fully populated, although because of the limited coupling between masses in Fig. 8, the mass matrix is diagonal and the damping and stiffness matrices are sparcely populated.

3.2 Modal Decomposition

Undamped System If the system is undamped and a harmonic force of the form $\{f\}e^{j\omega t}$ is applied to the system, the response can be readily calculated by assuming a response of the form $\{u\}e^{j\omega t}$ and simply inverting the resulting matrix to obtain

$$\{u\} = ([K] - \omega^2[M])^{-1}\{f\}, \qquad (27)$$

where $[A]^{-1}$ means the inverse of the matrix $[A]$ defined by $[A][A]^{-1} = [I]$, where $[I]$ is the unit diagonal matrix.[3] The response of the system to nonharmonic forcing can then be obtained using the Fourier or Laplace transformation techniques described in the previous section. While this approach is certainly possible, it provides little insight into the dynamics of the system. A useful alternate approach is to decompose the system into normal modes.

If a solution of the form $\{u\}e^{j\omega t}$ is assumed for the undamped system, one can write

$$\{f\} = ([K] - \omega^2[M])\{u\}. \qquad (28)$$

When the forcing is zero, in order for this equation to have a solution it is necessary that $\det([K] - \omega^2[M]) = 0$, where $\det([A])$ means the determinant of the matrix $[A]$. Setting the determinant to zero results in a polynomial equation in ω of the form

$$a_{2N}\omega^{2N} + a_{2N-2}\omega^{2N-2} + \cdots + a_0 = 0, \qquad (29)$$

which has N real solutions of the form $\pm\omega_i$ where N is the number of degrees of freedom of the system. These natural frequencies or eigenvalues of the system are analogous to the natural frequency of the single-degree-of-freedom system in Eq. (1). Sometimes not all of the solutions to Eq. (29) are different. For this so-called degenerate case the reader is referred to any of the standard texts on matrix algebra.

If one substitutes the natural frequencies from Eq. (29) into the system equations, Eq. (28), with zero external force, and solves for the amplitudes $\{u\}$, one obtains one mode shape $\{\phi_i\}$ (or eigenvector) for each natural frequency ω_i. It can be easily shown[3,4] that if $[K]$ and $[M]$

are real and symmetric, these mode shapes are composed of all real elements and are orthogonal to one another with respect to the mass and stiffness matrices, that is,

$$\begin{aligned} \{\phi_i\}^{\mathrm{T}}[M]\{\phi_j\} &= 0 \quad \text{for } i \neq j, \\ \{\phi_i\}^{\mathrm{T}}[K]\{\phi_j\} &= 0 \quad \text{for } i \neq j, \end{aligned} \tag{30}$$

where $\{\phi\}^{\mathrm{T}}$ means the transpose of the vector $\{\phi\}$. The mode shapes defined as described above are not unique but are defined only to within an arbitrary scale factor. One commonly used scaling factor is to set the maximum value in any mode shape to unity. With that scaling for the mode shapes the generalized mass and stiffness for mode i are defined as

$$\begin{aligned} \{\phi_i\}^{\mathrm{T}}[M]\{\phi_i\} &= m_i, \\ \{\phi_i\}^{\mathrm{T}}[K]\{\phi_i\} &= k_i \end{aligned} \tag{31}$$

and are related to the natural frequency of mode i by

$$\omega_i = \left(\frac{k_i}{m_i}\right)^{1/2}. \tag{32}$$

If one now combines the mode shapes defined above into a matrix such that each modal column vector $\{\phi_i\}$ forms a column of the matrix, that is, $[\Phi] = [\{\phi_1\}, \{\phi_2\} \cdots \{\phi_N\}]$, it is straightforward to show that

$$\begin{aligned} [\Phi]^{\mathrm{T}}[M][\Phi] &= [m_i], \\ [\Phi]^{\mathrm{T}}[K][\Phi] &= [k_i], \end{aligned} \tag{33}$$

where $[m_i]$ and $[k_i]$ are diagonal matrices. If one assumes a solution to Eq. (26) of the form

$$\{u\} = [\Phi]\{a(t)\}, \tag{34}$$

where the $a(t)$ are the modal amplitudes, substitutes this solution into the Eq. (26) with $[B] = 0$, and premultiplies the equation by $[\Phi]^T$, one obtains

$$[m_i]\left\{\frac{d^2a}{dt^2}\right\} + [k_i]\{a\} = [\Phi]^{\mathrm{T}}\{f\}. \tag{35}$$

Since the matrices $[m_i]$ and $[k_i]$ are diagonal, the above represents N uncoupled equations for N independent oscillators. Each of these equations can be solved independently of the other equations by any of the techniques described in the previous section. In the case where the external forces are all harmonic forces of the form $\{f\} = \{F\}e^{j\omega t}$ the solution $\{a\} = \{A\}e^{j\omega t}$ simplifies to

$$A_i = \frac{F_i}{m_i[\omega_i^2 - \omega^2]}, \tag{36}$$

where $F_i = \{\phi_i\}^{\mathrm{T}}\{F\}$ is the modal force for mode i and A_i is the ith element of $\{A\}$.

From this result and Eq. (34) the vector of displacement amplitudes of the individual masses $\{U\}$ becomes

$$\{U\} = [\Phi]\left[\frac{1}{m_i(\omega_i^2 - \omega^2)}\right][\Phi]^{\mathrm{T}}\{F\}, \tag{37}$$

where $[1/m_i(\omega_i^2 - \omega^2]$ is a diagonal matrix.

The matrix $[\Phi][1/m_i(\omega_i^2 - \omega^2)][\Phi]^{\mathrm{T}}$ is the multimodal reactance matrix. It is analogous to Eq. (15), the reactance equation for a single-degree-of-freedom system. Now instead of a single frequency response function there are N^2 of such functions, although because of symmetry there are only $(N^2 - N)/2 + N$ different ones. Similar equations can be written for the admittance and accelerance matrices and the inverses of these quantities, for example, apparent stiffness, apparent mass, and impedance, can be obtained by simply taking the matrix inverse. For example, the admittance matrix $[Y]$ can be obtained from

$$\{V\} = j\omega\{U\} = [\Phi]\left[\frac{j\omega}{m_i(\omega_i^2 - \omega^2)}\right][\Phi]^{\mathrm{T}}\{F\} \tag{38}$$

and

$$[Y] = [\Phi]\left[\frac{j\omega}{m_i(\omega_i^2 - \omega^2)}\right][\Phi]^{\mathrm{T}}.$$

The impedance matrix $[Z]$ is given by

$$[Z] = [Y]^{-1}.$$

Damped System With the inclusion of damping modal decomposition is still possible but in the general case considerably more complicated. One simplified damping model commonly used in conjunction with viscous damping leads to modal decomposition that is closely related to the undamped case. This simplified model is called *proportional* damping. With proportional damping the damping matrix is given by

$$[B] = \alpha[M] + \beta[K], \tag{39}$$

where α and β are constants. In this instance the modal decomposition is simply related to the undamped system. However, now the eigenvalues are complex and occur in pairs given by

$$\omega_i' = \omega_i[j\zeta_i \pm (1 - \zeta_i^2)^{1/2}],$$
$$\zeta_i = \frac{\alpha}{2\omega_i} + \frac{\beta\omega_i}{2}, \tag{40}$$

where ω_i is the undamped natural frequency of mode i. The mode shapes, however, are real and are the same as the undamped system mode shapes

$$\{\phi_i^{\text{damped}}\} = \{\phi_i^{\text{umdamped}}\}.$$

In this case the damped system modal decomposition and the undamped system modal decomposition are simply related. Unfortunately, the requirement that the damping matrix be a linear combination of the mass and stiffness matrices is very restrictive, and more general models of damping are often necessary.

Proportional damping can also be applied to the case of hysteretic damping. However, the general case of hysteretic damping is nearly as simple. For harmonic excitation of a system with hysteretic damping, the vectors of force amplitudes $\{f\}$ and response amplitudes $\{u\}$ are related by

$$(-\omega^2[M] + [K] + j[H])\{u\} = \{f\}, \tag{41}$$

where the system response has been assumed to be of the form $\{u\}e^{j\omega t}$. Proceeding as for the undamped case, one can define a set of N eigenvalues $\pm\omega_i'$ (N is the number of degrees of freedom of the system) by solving

$$\det(-\omega^2[M] + [K] + j[H]) = 0. \tag{42}$$

An eigenvector ϕ_i' can be obtained for each eigenvalue by substituting each ω_i' into Eq. (41), solving for $\{u\}$ with the force vector $\{f\}$ set to zero and normalizing the result as for the undamped case. In this instance, however, because the system matrix is complex and not real, the resulting eigenvalues and eigenvectors will be complex. Similar orthogonality conditions apply as for the undamped case, that is,

$$[\Phi']^{\mathrm{T}}[M][\Phi'] = [m_i'],$$
$$[\Phi']^{\mathrm{T}}[K][\Phi'] = [k_i'], \tag{43}$$

where $[\Phi']$ is the matrix whose columns are formed by the eigenvectors (mode shapes) $\{\phi_i\}$ and $[m_i']$ and $[k_i']$ are diagonal matrices. In addition it can be shown that

$$(\omega_i')^2 = \frac{k_i'}{m_i'}$$

where k_i' and m_i' are the diagonal elements of the matrices $[k_i']$ and $[m_i']$, respectively. The response can be estimated using the same equations as for the undamped case [Eq. (37)] with the complex eigenvalues ω_i' and complex eigenvectors $\{\phi_i'\}$ substituted for their real counterparts in the undamped equations. It should be noted that the complex mode shapes or eigenvectors simply allow for a general variation in phase between elements of the eigenvectors as opposed to requiring that every element be either in phase or out of phase as is the case in the undamped system.

For the case of viscous damping that is not proportional a slightly different modal decomposition procedure can be applied to obtain a set of $2N$ eigenvalues consisting of N complex eigenvalues and their complex conjugates. There will also be $2N$ eigenvectors consisting of N complex eigenvectors and their N complex conjugates. However, for estimating the forced response, it is more convenient to reformulate the equations and put them in standard state-space form. If a new vector, $\{p\}$, $2N$ elements long consisting of the displacements of each mass and its velocity is defined as

$$\{p\} = \begin{bmatrix} \{u\} \\ \left\{\dfrac{du}{dt}\right\} \end{bmatrix}, \tag{44}$$

then the equations of motion can be rewritten in the form

$$[R]\left\{\frac{dp}{dt}\right\} + [S]\{p\} = \{q\}, \tag{45}$$

where

$$[R] = \begin{bmatrix} [B] & [M] \\ [M] & [0] \end{bmatrix}, \qquad [S] = \begin{bmatrix} [K] & [0] \\ [0] & -[M] \end{bmatrix},$$
$$\{q\} = \begin{bmatrix} \{f\} \\ \{0\} \end{bmatrix}.$$

These equations can be modally decomposed into $2N$ complex eigenvalues ω_i' (N complex-conjugate pairs) and $2N$ eigenvectors $\{\phi_i'\}$ (N complex-conjugate pairs) of length $2N$. The eigenvectors possess orthogonality properties of the form

$$[\Phi']^{\mathrm{T}}[R][\Phi'] = [r_i],$$
$$[\Phi']^{\mathrm{T}}[S][\Phi'] = [s_i], \tag{46}$$

where $[\Phi]$ is a $2N \times 2N$ matrix the columns of which are the eigenvectors $\{\phi_i'\}$ and $\omega_i' = s_i/r_i$. As was found for the undamped case, if a solution of the form $\{p\} = [\Phi']\{a\}$ is substituted into Eq. (45), one obtains the solution

$$\{p\} = [\Phi'][A][\Phi']^{\mathrm{T}}\{q\}, \tag{47}$$

where $[A]$ is a diagonal matrix with elements given by

$$A_{ii} = [r_i(j\omega + \omega_i')]^{-1}.$$

Although this form appears quite different form previous cases, the fact that the eigenvalues and eigenvectors come in complex conjugate pairs allows Eq. (47), when it is expanded, to take an appearance similar to the undamped case.

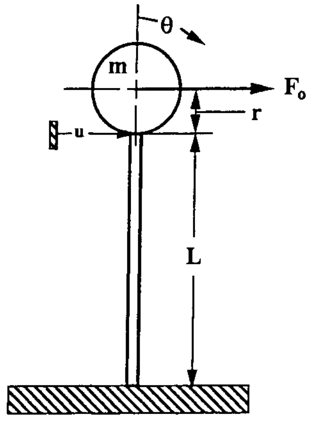


Fig. 9 Mass with significant rotary inertia on the end of a cantilever beam.

Example As an example we will reexamine the two-degree-of-freedom system illustrated in Fig. 9. This system is similar to the single-degree-of-freedom system already analyzed in Fig. 4. Here, we will allow the mass at the end of the cantilever beam to have significant rotary inertia. If the beam is assumed to deform quasi-statically, the matrix equation relating force and moment at the end of the beam to the resulting deflection and rotation is given by

$$\begin{bmatrix} F_b \\ M_b \end{bmatrix} = \begin{bmatrix} \dfrac{12EI}{L^3} & \dfrac{-6EI}{L^2} \\ \dfrac{-6EI}{L^2} & \dfrac{4EI}{L} \end{bmatrix} \begin{bmatrix} u \\ \theta \end{bmatrix}, \tag{48}$$

where EI is the bending stiffness of the beam, θ is the angle of rotation of the mass about its center of gravity, u is the lateral displacement of the end of the beam where it joins the mass, F_b is the force and M_b is the moment at the interface between the mass and the end of the beam, and the other variables are as defined previously. A similar set of equations can be written for the mass

$$\begin{bmatrix} F_b \\ M_b \end{bmatrix} = \begin{bmatrix} -m & -mr \\ -mr & I_m + mr^2 \end{bmatrix} \begin{bmatrix} \dfrac{d^2u}{dt^2} \\ \dfrac{d^2\theta}{dt^2} \end{bmatrix} + \begin{bmatrix} F_0 \\ -F_0 r \end{bmatrix}, \tag{49}$$

where F_0 is the force applied to the center of mass of the mass and I_m is the rotary inertia of the mass. Combining the two equations, we obtain the equation of motion of the system

$$\begin{bmatrix} m & mr \\ mr & I_m + mr^2 \end{bmatrix} \begin{bmatrix} \dfrac{d^2u}{dt^2} \\ \dfrac{d^2\theta}{dt^2} \end{bmatrix} + \begin{bmatrix} \dfrac{12EI}{L^3} & \dfrac{-6EI}{L^2} \\ \dfrac{-6EI}{L^2} & \dfrac{4EI}{L} \end{bmatrix} \times \begin{bmatrix} u \\ \theta \end{bmatrix} = \begin{bmatrix} F_0 \\ -F_0 r \end{bmatrix}. \tag{50}$$

As the equation clearly shows, the system has two degrees of freedom. To illustrate the solution of this system to harmonic excitation, consider solutions of the form $e^{j\omega t}$ for u and θ. The resulting equation becomes

$$\left\{ -\omega^2 \begin{bmatrix} 1 & r \\ r & \dfrac{I_m}{m} + r^2 \end{bmatrix} + \omega_0^2(1 + j\eta) \begin{bmatrix} 1 & -\dfrac{L}{2} \\ -\dfrac{L}{2} & \dfrac{L^2}{3} \end{bmatrix} \right\} \cdot \begin{bmatrix} u \\ \theta \end{bmatrix} = \begin{bmatrix} \dfrac{F_0}{m} \\ -\dfrac{F_0 r}{m} \end{bmatrix}, \tag{51}$$

where $\omega_0^2 = 12EI/mL^3$ and where hysteretic damping proportional to the stiffness terms in the equation has been assumed. This form of damping is commonly used if the modulus of the beam can be modeled as complex and of the form $E = E(1 + j\eta)$. If the physical properties of the system are given by

$$r = 0.001 \text{ m}, \qquad L = 0.1 \text{ m}, \qquad m = 0.1 \text{ kg},$$

$$EI = 0.0001 \text{ N} \cdot \text{m}^2, \qquad \frac{I_m}{m} = 0.0001 \text{ m}^2, \qquad \eta = 0.05.$$

Setting the determinant of the matrices in the brackets in Eq. (51) to zero and solving for the two natural frequencies result in

$$\begin{aligned} f_1 &= \frac{\omega_1}{2\pi} = 0.27 + 0.0067j \quad \text{Hz}, \\ f_2 &= \frac{\omega_2}{2\pi} = 3.27 + 0.082j \quad \text{Hz}. \end{aligned} \tag{52}$$

The natural frequencies are complex due to the damping, with the real parts only slightly different from the real natural frequencies that would result from the undamped system. Substituting these frequencies back into Eq. (51) and solving for θ in terms of u, one obtains the modal vectors (eigenvectors) associated with each natural frequency (eigenvalue):

$$\phi_1 = \begin{bmatrix} 0.0657 \\ 0.998 \end{bmatrix}, \qquad \phi_2 = \begin{bmatrix} -0.0025 \\ 1.000 \end{bmatrix}. \tag{53}$$

These modal vectors are real and the same as one would obtain from the undamped system because the damping matrix is proportional to the stiffness matrix. The system response to a harmonic force F_0 can be obtained by substituting the above results into Eqs. (33) and (36), which results in

$$\begin{bmatrix} \frac{u}{F_0} \\ \frac{\theta}{F_0} \end{bmatrix} = [\Phi] \begin{bmatrix} \frac{1}{m_1(\omega_1^2 - \omega^2)} & 0 \\ 0 & \frac{1}{m_2(\omega_2^2 - \omega^2)} \end{bmatrix} \cdot [\Phi]^{\mathrm{T}} \begin{bmatrix} 1 \\ -r \end{bmatrix},$$

where the modal masses ($m_1 = 4.55 \times 10^{-4}$, $m_2 = 1.02 \times 10^{-5}$) are computed using Eq. (33). The result of carrying out the indicated matrix multiplications is

$$\begin{aligned} \frac{u}{F_0} &= \frac{9.3}{\omega_1^2 - \omega^2} + \frac{0.86}{\omega_2^2 - \omega^2} \quad \text{m/N}, \\ \frac{\theta}{F_0} &= \frac{142}{\omega_1^2 - \omega^2} - \frac{343}{\omega_2^2 - \omega^2} \quad \text{rad/N}, \end{aligned} \tag{54}$$

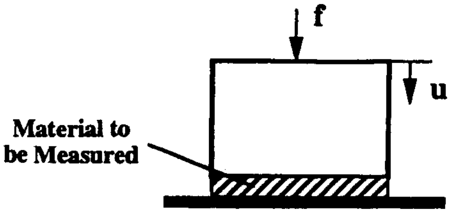
where ω_1 and ω_2 are given by Eq. (52).

4 EXPERIMENTAL METHODS

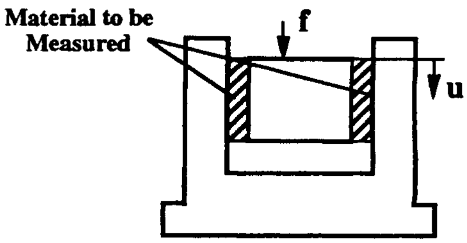
4.1 Material Property Estimation Using Simple Oscillators

A common approach to the measurement of material properties is the use of *resonance testing*, so called because it typically involves the coupling of a known mass with the unknown material, and the response of the mass on the material near resonance is used to infer the stiffness and damping properties. A wide variety of arrangements of mass and material are commonly used for this type of testing. For example, in Fig. 10 two arrangements are shown, one appropriate for measuring the compressional properties of a pad of unknown material and one appropriate for measuring the shear properties. (See also Chapter 72.)

One also has a number of choices in how the sample is excited and in what quantities are measured. Figure 11 shows a schematic picture of a typical arrangement in which the pad is mounted on a rigid platform, the mass is mounted on top of the pad, a known force is applied to the mass, and the vertical motion of the mass is measured. The force is usually applied by an electromagnetic shaker as shown in the figure, and the force and acceleration are measured using an impedance head or force transducer and accelerometer. The influence of static load on the material properties is often of interest as well. The figure shows the commonly used approach for apply-



(a) Compressive Stiffness



(b) Shear Stiffness

Fig. 10 Alternate arrangements for material property testing.

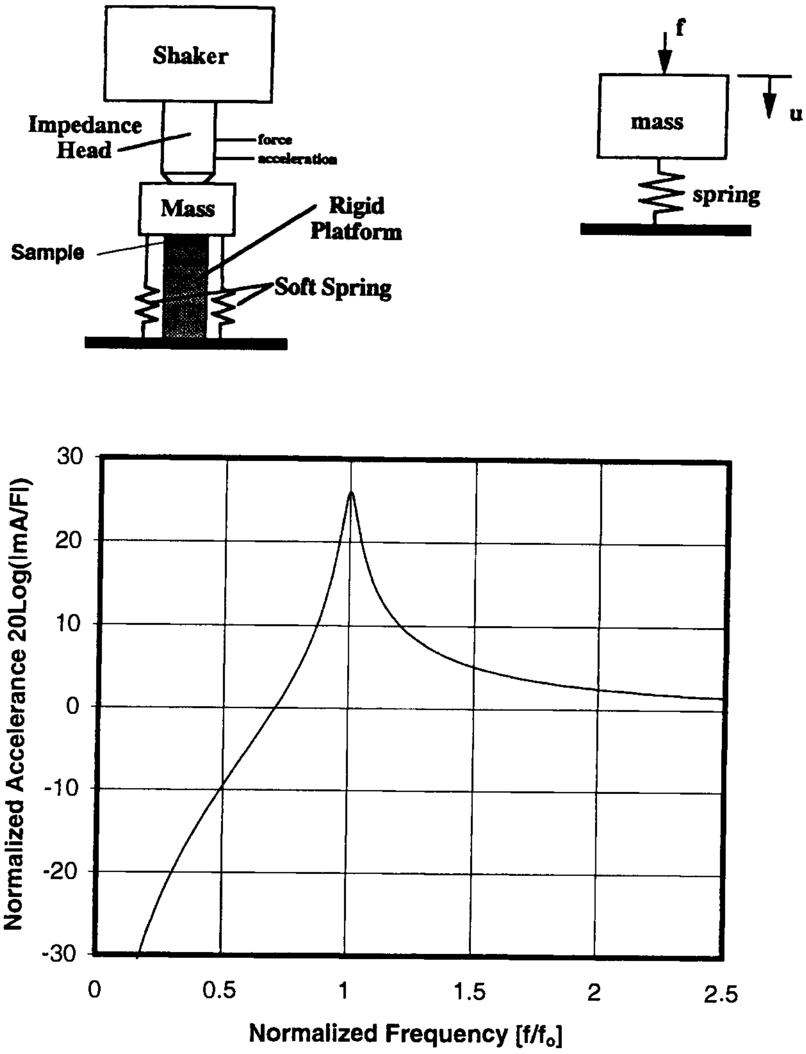


Fig. 11 Typical material property test arrangement including static loading of the sample.

ing a static load in which the load is applied through a spring with a stiffness much less than the pad stiffness. Properly designed such an arrangement will not affect the measured stiffness significantly. In practice several accelerometers are placed on the mass to ensure that the motion is purely vertical translation.

For the arrangement shown in Fig. 11, the ratio of acceleration to applied force (accelerance) is given by

$$\frac{A}{F} = \frac{\omega^2\{(k - \omega^2 m) - j\eta k\}}{\{(k - \omega^2 m)^2 + [\eta k]^2\}}, \tag{55}$$

where a hysteretic damping model for the pad has been assumed, that is, $k = k(1+j\eta)$. A plot of the amplitude of this transfer function for $\eta = 0.05$ is given at the bottom of Fig. 11. Given measured data of the form of Fig. 11, it is the intent to estimate the stiffness k and the loss factor η. Theoretically, k and η can be obtained at any frequency by simply measuring the amplitude and phase of a/F and appropriately solving Eq. (55) for η and k. In practice, there are a number of approaches that make the process easier and less prone to error.

For estimating k the following approaches are typically applied:

- For $\omega \ll (k/m)^{1/2}$ the accelerance is dominated by the pad stiffness and

$$\left|\frac{A}{F}\right| \sim \frac{\omega^2}{k}. \tag{56}$$

From this equation k can be found at any frequency for which the above criterion holds.

- The $\mathrm{Re}\{A/F\} = 0$ when

$$\omega = \left(\frac{k}{m}\right)^{1/2}. \tag{57}$$

Consequently, if a multichannel fast Fourier transform (FFT) analyzer is available, one can excite the mass with band-limited white noise, use the analyzer to compute the $\mathrm{Re}\{A/F\}$ as a function of frequency, and search for the frequency at which it goes to zero. In addition when the $\mathrm{Re}\{A/F\} = 0$, the phase between A and F is 90°. Therefore this frequency can also be found by applying a harmonic force to the mass, varying the frequency and monitoring the phase between the force and acceleration until that phase becomes 90°. Once the frequency at which $\mathrm{Re}\{A/F\} = 0$ (phase of $A/F = 90°$), the stiffness at that frequency can be found from the above equation. To obtain k at other frequencies, one must use a different mass and retest, using lighter masses for higher frequencies and heavier masses for lower frequencies.

- The amplitude $|A/F|$ peaks at

$$\omega_p = \left\{\frac{k}{m[1+\eta^2]}\right\}^{1/2}. \tag{58}$$

Consequently, if the damping is very small, the dynamic stiffness at ω_p can be estimated approximately by

$$k \sim \omega_p^2 m. \tag{59}$$

As before the dynamic stiffness at other frequencies can be obtained by changing the mass and retesting the sample.

The following are a few examples of typical approaches for estimating the loss factor:

- At the frequency where $\mathrm{Re}\{A/F\} = 0$ and the phase between A and F is 90° [$\omega = (k/m)^{1/2}$], the amplitude of A/F becomes

$$\left|\frac{A}{F}\right| = \frac{\omega^2}{k\eta} \tag{60}$$

from which η can be readily calculated once k has been determined. To obtain estimates of η at other frequencies, the mass must be changed as before and the sample retested.

- The half-power bandwidth at resonance offers another technique for estimating the damping. If the frequency at which $|A/F|$ peaks is given by ω_p, then there is a frequency ω_h above ω_p and a frequency ω_l below ω_p at which $20 \log(|A/F|)$ decreases from its maximum by 3 dB. If the damping is small, $\eta \ll 1$, then

$$\eta \approx \frac{\omega_h - \omega_l}{\omega_p}. \tag{61}$$

Once again to obtain η at other frequencies, the mass must be changed to change ω_p and the sample must be retested.

Two of these techniques have been used to estimate the mechanical properties of a number of cork and glass fiber pads for which measured accelerance data with a mass on the pads under two different static loads were available.[5] These data were obtained using an apparatus similar to that shown in Fig. 11. The dynamic stiffness was estimated using direct measurement well below the resonance frequency [Eq. (56)] and resonance testing [Eq. (59)] when η is small. As Table 2 shows, the two approaches give comparable results.

4.2 Experimental Modal Analysis

The measurement and data analysis techniques available for estimating the natural frequencies, damping, and mode shapes of multimodal systems are commonly referred to as experimental modal analysis.[4–6] While applicable to the experimental analysis of any multimodal dynamic system, these techniques have been most commonly applied to structures.

Experimental modal analysis has been the subject of considerable attention in recent years. This interest was initially spurred by the introduction of fast, increasingly capable, multichannel FFT analyzers, many of which came packaged with modal analysis software, and by the availability of associated instrumentation (impedance

TABLE 2 Dynamic Stiffness Test Results on 5-cm Cubes of Material

Test Sample	Resonance Testing (N/m)	Direct Stiffness (N/m)
Cork		
High density	9.71×10^5	9.79×10^5
Standard density	4.17×10^5	4.92×10^5
Low density	3.17×10^5	3.14×10^5
Glass fiber pad	1.30×10^5	1.23×10^5

Source: Ref. 5. Reprinted with permission of the author.

hammers, etc.). Both of these developments eased the tedious process of acquiring the necessary transfer function data. Further development and use of experimental modal analysis techniques were encouraged by the availability of increasingly less expensive personal computers and workstations that provided greater flexibility to the analyst in the processing of the transfer function data.

Experimental modal analysis consists of two major steps:

- The acquisition of measured transfer function data relating point force excitation at one or more points on a structure to response at a number of other points
- The processing of the transfer function data to estimate the natural frequencies, damping factors, and mode shapes

Experimental modal analysis may be carried out using either frequency-domain or time-domain techniques. The earliest procedures, utilizing frequency domain techniques, involved the used of arrays of shakers operating at a single frequency. The shakers were tuned to exactly balance the dissipative focus in the structure and thereby allow the estimation of the homogeneous, undamped equations of motion of the structure.[7] These approaches eventually evolved into the frequency response function (FRF) methods that are most commonly used today.[8] Time-domain techniques utilize the transient response characteristics of the structure to obtain the modal parameters. Typical examples of time-domain techniques would be the Ibrahim time-domain method[9] and the polyreference time-domain method.[10] All of these techniques are in current use and new approaches are constantly evolving.[11]

Data Acquisition The goal of the FRF method data acquisition program is to obtain good-quality frequency response functions relating the force applied to a number of points on the structure to the response (displacement, velocity, or acceleration) at a number of points on the structure, where as least some of the force and response points are the same. As with any carefully designed experimental program, experimental modal analysis must begin with the proper selection of transducers, signal conditioning electronics, and analysis hardware. Since these issues are not unique to modal analysis and are discussed adequately elsewhere in this handbook, they will be dealt with no further in this section.

A serious consideration is the mounting of the structure to be analyzed. Two types of mountings are often considered: free and fixed. In a free mounting the structure is suspended with very soft attachments. The attachments are generally designed to be soft enough, such that any resonances of the mass of the structure (in any of its rigid body degrees of freedom) on the suspension are at frequencies well below the frequency range of interest for the measurements. In the fixed mounting the structure is firmly attached to a *rigid* foundation, designed to be essentially motionless compared to the structure during the test. It is generally much easier to design a free suspension than to design one that is fixed. If the suspension is not properly designed, that is, free or fixed, then, the dynamics of the mounting system may contaminate the modal test results.

Once the structure has been properly mounted, a grid work of measurement points must be laid out on it. This is typically accomplished by carrying out a series of preliminary transfer function measurements at a number of points on the structure. These measurements allow the experimentalist to select the excitation levels to obtain good coherence in the measured transfer functions. Coherence is a data processing function available on most multichannel FFT analyzers that provides information on the quality of the data used to form a transfer function, for example, signal-to-noise ratio and linearity of the signals. The measurements also help to establish the required spacing between measurement points in the measurement grid. If a finite-element model of the structure is available, then, the initial definition of the measurement grid can be based on the predicted mode shapes.

Once the measurement grid has been laid out, the measurements may follow one of two possible forms. The experimenter may fix the location where the response is measured and apply the desired force to each of the grid points, or he may fix the point of application of the force and measure the response at each of the grid points. Usually more than one fixed response point or more than one fixed force point is used in order to minimize the likelihood that the fixed point will be a node of one of the modes of interest, that is, a point where the modal response is zero.

A variety of forms of excitation are used to excite the structure. If an electrodynamic shaker or similar device in conjunction with a force gauge is used, the excitation is usually swept sine, stepped sign, pseudo-random noise (digitally generated random noise), or random noise. For impulsive excitation a device usually referred to as an impedance hammer is used. The impedance hammer is a hammer with a force gauge fitted to the head. It is especially convenient for modal analysis measurements in which a few fixed response points are used with the excitation force applied to all of the grid points.

Data Analysis The analysis of the FRF data is usually but not always split into two parts: (1) the estimation of the natural frequencies, modal damping, and modal coefficients and (2) the estimation of the mode shape functions.

Software for performing both of these functions is generally available for PCs, workstations, and mainframe computers and in some cases comes bundled with multichannel FFT analyzers.

In principle, the first step requires only one FRF. In practice, however, because the point of force application or the point of response measurement may be at a node of one or more modes of the structure, data from most if not all of the FRFs are normally utilized.

In the simplest case of lightly damped, well-separated modes, the natural frequencies can be determined from a simple inspection of the FRF curve, and the modal damping coefficients can be estimated based on the half-power bandwidth of each resonance divided by the resonant frequency [see Eq. (61)]. The modal constants $(A_{kl})_q$ can be obtained by noting that

$$Y_{kl}(\omega_a) = \sum_{q=n1}^{N+ni} \frac{(A_{kl})_q}{\omega_q^2 - \omega_a^2}, \tag{62}$$

where $Y_{kl}(\omega_a)$ is a frequency response function at frequency ω_a relating the response measured at point k to the force applied at point l. The summation over q is over the modes of the structure. By forming an equation of the form of Eq. (62) for each frequency where measured data are available, one can form a matrix equation of the form

$$\begin{bmatrix} Y_{kl}(\omega_a) \\ Y_{kl}(\omega_b) \\ \vdots \end{bmatrix} = \begin{bmatrix} (\omega_1^2 - \omega_a^2)^{-1} & (\omega_2^2 - \omega_a^2)^{-1} & \cdots \\ (\omega_1^2 - \omega_b^2)^{-1} & (\omega_2^2 - \omega_b^2)^{-1} & \cdots \\ \vdots & \vdots & \cdots \end{bmatrix} \cdot \begin{bmatrix} (A_{kl})_1 \\ (A_{kl})_2 \\ \vdots \end{bmatrix}$$

or

$$\{Y_{kl}\} = [B]\{A_{kl}\}, \tag{63}$$

in which the matrix $[B]$ can be readily inverted to obtain the unknown modal constants in terms of the measured FRFs $\{Y_{kl}\}$ provided the number of data points is at least as great as the number of modal coefficients to be determined.

In practice, however, overlap between modes can distort the FRF sufficiently so as to obscure the response of individual modes. In such a case more sophisticated approaches must be employed. One of the original approaches to obtaining a simultaneous estimate of all modal parameters is based on a least-mean-square curve-fitting procedure.[12] While more sophisticated procedures are available today, this approach is still useful especially when computational resources are limited, and it is worthwhile spending a moment examining it. One begins with a set of FRFs, $Y_{kl}^{(m)}(\omega_i)$, measured at a number of frequencies ω_i. If the FRFs are receptances, and if we choose to use a structural damping model, the mathematical model to be fitted to the measured data is given by

$$Y_{kl}^{(e)}(\omega_i) = \sum_{q=n1}^{N+ni} \frac{(A_{kl})_q}{\omega_q^2 - \omega_i^2 + j\eta_q\omega_q^2} + \frac{1}{k_{kl}} - \frac{1}{\omega_i^2 m_{kl}}, \tag{64}$$

where the summation is the usual one over N modes, and the final two terms account for the influence of modes not included in the summation. The intent is to determine the modal parameters $\omega_{n1}, \omega_{n1+1}, \ldots, \eta_{n1}, \eta_{n1+1}, \ldots, (A_{kl})_{n1}, (A_{kl})_{n1+1}, \ldots, k_{kl}, m_{kl}$ so that the equation fits the measured data points with the minimum mean-square error. In mathematical terms, if the mean-square error is given by

$$e^2 = \sum_i |Y_{kl}^{(e)}(\omega_i) - Y_{kl}^{(m)}(\omega_i)|^2, \tag{65}$$

then it is desired to select the above-mentioned modal parameters so as to minimize e^2. To estimate the parameters, one differentiates Eq. (65) with respect to each modal parameter and sets the resulting equations equal to zero,

$$\frac{d(e^2)}{db} = 0, \tag{66}$$

where

$$b = \omega_{n1}, \omega_{n1+1}, \ldots, \eta_{n1}, \eta_{n1+1}, \ldots, (A_{kl})_{n1}, (A_{kl})_{n1+1}, \ldots, k_{kl}, m_{kl}.$$

Unfortunately Eq. (66) results in $4N + 2$ nonlinear simultaneous equations [the modal constants $(A_{kl})_q$ are complex and hence two parameters are required for each] in the unknown parameters. Such a set of equations cannot, in general, be solved directly. As a result iterative approaches (typically Newton–Raphson procedures) have been developed in which an initial estimate of the desired parameters must be provided. The initial estimate may be based on inspection of the real and imaginary parts of the FRF or more sophisticated measures may be employed.[12]

In the time domain a comparable procedure exists[4] that is called the complex exponential method. To apply it, one must have a time series representing the impulse response

$$h_{kl}(\Delta t), h_{kl}(2\ \Delta t), \ldots, h_{kl}(p\ \Delta t),$$

where, as before, l refers to the point of force application and k is the response point. It is desired to fit these data with a mathematical model of the form

$$h_{kl}(t) = \sum_{q=1}^{N} (A_{kl})_q e^{s_q t} + (A_{kl})_q^* e^{s_q^* t},$$

$$s_q = \omega_q[-\zeta_q + j(1 - \zeta_q^2)^{1/2}].$$

Using this equation at each time point, one obtains

$$\begin{bmatrix} h_{kl}(0) \\ h_{kl}(\Delta t) \\ h_{kl}(2\ \Delta t) \\ h_{kl}(3\ \Delta t) \\ \vdots \\ h_{kl}(p\ \Delta t) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & & 1 \\ V_1 & V_1^* & V_2 & \cdots & V_N^* \\ V_1^2 & (V_1^*)^2 & V_2^2 & \cdots & (V_N^*)^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ V_1^p & & & & (V_N^*)^p \end{bmatrix} \cdot \begin{bmatrix} (A_{kl})_1 \\ (A_{kl})_1^* \\ (A_{kl})_2 \\ \vdots \\ (A_{kl})_N^* \end{bmatrix}, \tag{67}$$

where $V_q = e^{s_q \Delta t}$.

One can solve Eq. (67) for V_1 to V_N using the Prony method[13] provided the number of data points exceeds $4N$. From the V's one obtains ω_i and ζ_i, and by substituting back into Eq. (67) and inverting the matrix of V's, one can obtain the modal constants $(A_{kl})_i$.

The parameter estimation procedures described so far operate on a single FRF or impulse response function (IRF). In practice, the procedures are applied to a number of different FRF/IRFs, and some average over the resulting modal parameters is used to obtain the desired estimates. Today, more sophisticated procedures are available that simultaneously fit a number of FRF/IRFs at once.[9,10] In fact the Ibrahim time-domain method[10] not only accepts all IRFs at once, it automatically estimates not only the modal frequencies and damping but the mode shape functions as well.

An example of applying a multi-degree-of-freedom curve-fitting process to the measurement of mobility on a substructure of a helicopter is shown in Fig. 12.[4] In the figure the modes were well separated enough and the damping light enough that the process for well-separated lightly damped modes could be applied quite successfully.

Once the natural frequencies, modal damping, and modal constants have been obtained by the curve-fitting process, the next step is the derivation of the mode shape functions. If the measured FRFs are receptances, then the relationship between modal constant and mode shape function is given by

$$(A_{kl})_i = \frac{\phi_i(x_k)\phi_i(x_l)}{m_i}. \tag{68}$$

For those measurements in which the force application point and the response measurement point are co-located, the relationship becomes

$$(A_{kk})_i = \frac{\phi_i(x_k)^2}{m_i}. \tag{69}$$

With Eqs. (68) and (69), one can solve for the mode shape functions $\phi_i(x)$.

Once the modal parameters have been estimated, they can be used in two general ways:

- To validate numerical models of the structure (e.g., finite element) through comparison of predicted and measured natural frequencies and mode shapes
- To generate a mathematical model of the structure

Once a mathematical model has been generated from a modal analysis test, there are a variety of uses to which it can be put. A few examples follow:

- To model a subassembly of a larger structure for integration into the total structural model
- To estimate the effects of structural modifications
- To estimate the excitation force applied to a structure when direct measurement of the force is difficult

In short, experimental modal analysis is a powerful tool that can be used alone or in conjunction with analytical models of structures to better understand the dynamics of complex structures and to aid in their design to ensure that they will function properly in demanding environments.

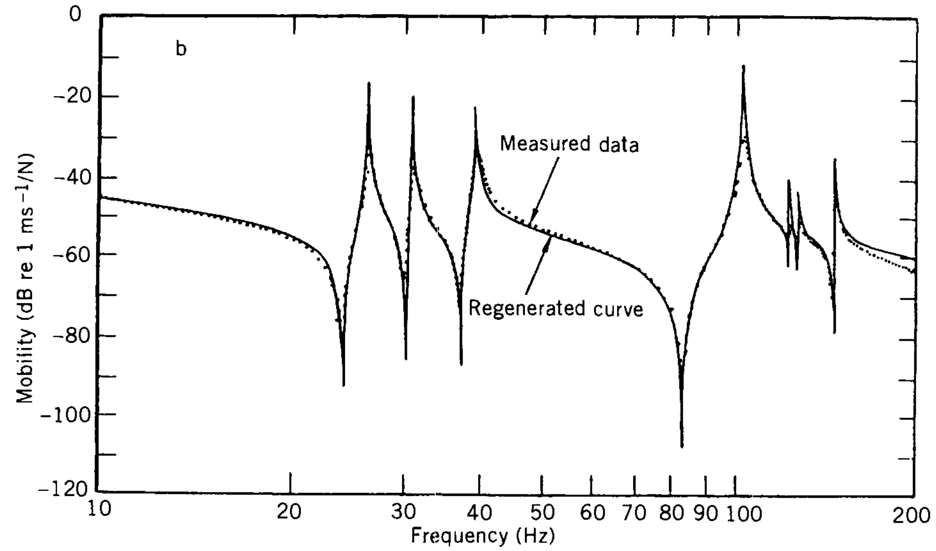


Fig. 12 Multi-degree-of-freedom curve fit to measured mobility on a helicopter auxiliary structure.[4] (Reprinted with permission of the author.)

REFERENCES

1. P. M. Morse, *Vibration and Sound*, McGraw-Hill, New York, 1948.
2. L. E. Kinsler and A. R. Frey, *Fundamentals of Acoustics*, Wiley, New York, 1962.
3. F. B. Hildebrand, *Advanced Calculus for Applications*, Prentice-Hall, Englewood Cliffs, NJ, 1976, Chap. 5.
4. D. J. Ewins, *Modal Testing Theory and Practice*, Wiley, New York, 1984.
5. I. L. Ver, "Measurement of Dynamic Stiffness and Loss Factor of Elastic Mounts as a Function of Frequency and Static Load," *Noise Control Eng.*, Vol. 3, No. 3, pp. 37–43, 1974.
6. R. J. Allemang, D. L. Brown, and R. W. Rost, "Experimental Modal Analysis and Aerodynamic Component Synthesis," Air Force Wright Aeronautical Laboratory Report No. 87-3069, 1987.
7. R. C. Lewis and D. L. Wristley, "A System for the Excitation of Pure Natural Modes of a Complex Structure," *J. Aeronaut. Sci.*, Vol. 17, No. 11, 1950, pp. 705–722.
8. D. L. Brown, R. D. Zimmerman, R. J. Allemang, and M. Mergeay, "Parameter Estimation Techniques for Modal Analysis," SAE Paper No. 790221, Society of Automotive Engineers Transactions, Vol. 88, 1979, pp. 828–846.
9. S. R. Ibrahim and E. C. Mikulcik, "A Method for the Direct Identification of Vibration Parameters for the Free Response," *Shock Vib. Bull*, Vol. 47, No. 4, 1977, pp. 183–198.
10. H. Vold and T. Rocklin, "The Numerical Implementation of a Multiple Input Modal Estimation Algorithm for Minicomputers," *Proceedings of the International Modal Analysis Conference (IMAC)*, Orlando, FL, 1982, pp. 542–548.
11. Excellent sources of information on all aspects of modal analysis are in the *Proceedings of the International Modal Analysis Conference (IMAC)* held annually since 1982 and sponsored by Union College, Schenectady, NY, 12308.
12. G. R. Gaukroger, C. W. Skingle, and K. H. Heron, "Numerical Analysis of Vector Response Loci," *J. Sound Vib.*, Vol. 29, No. 3, 1973, pp. 341–353.
13. R. Prony, "Essai experimental et analytique sur les lois de la delatabilite des fluides et sur celles de la force expansive del la vapeur de l'eau el de l'alkool differentes temperatures," *J. l'Ecole Polytech. (Paris)*, Vol. 1, No. 2, 1795, pp. 24–76.

64

VIBRATIONS OF ONE- AND TWO-DIMENSIONAL CONTINUOUS SYSTEMS

MANFRED A. HECKL

1 INTRODUCTION

Although all real structures are definitely three dimensional, the investigation of one- and two-dimensional systems always has been a field of active scientific interest.

The reasons for that are first that in the audio frequency range many structures such as beams, bars, columns, plates, membranes, shells, fuselages, ship hulls, pipes, and rings are "thin"; that is, at least in one dimension they are small compared to the relevant wavelength. Second, mathematically the treatment of structures in two dimensions is much simpler than in three dimensions.

In this text the important equations of motion are given. They follow from very basic principles of mechanics such as Newton's laws or Hamilton's principle and Hooke's law. The influence of damping usually is then "added on" in a somewhat heuristic way. From the equations of motion, wave speeds, different types of impedances, resonances, energy transmission quantities, impulse response functions, and so on, can be found. These quantities are very useful for noise control work, especially in vehicles and buildings.

Since all two-dimensional results are approximations, the limits of applicability are given whenever possible.

Throughout this chapter linearity is assumed; that is, material nonlinearities (e.g., in rubber) and geometric nonlinearities (e.g., when the amplitude of the motion is comparable with the thickness) are not taken into acount. (See Chapter 64.)

2 IMPORTANT RELATIONS FOR PLANAR SYSTEMS

In the following the most important parameters describing mechanical vibrations will be listed. Since there are several idealizations and approximations, depending on the geometry of the system and the direction of motion, only two cases—longitudinal motion in a bar and bending of a plate—will be discussed at some length in the main text. For other cases the relevant equations are given in the appendix.

2.1 Equations of Motion

Different types of wave motion are sketched in Fig. 1. They are approximations that are based on the assumption that the structure is sufficiently thin. In homogeneous media the wave types are independent; therefore they can be treated independently.

Longitudinal Waves in a Bar Longitudinal waves in a one-dimensional structure are of some importance for the vibration transmission over large distances in buildings, ships, and so on. Usually it is difficult to excite a bar or a column to pure longitudinal waves because their input impedance is quite high. As a consequence, slight unsymmetries cause additional bending waves that have a much lower input impedance. For sound radiation pure longitudinal waves are of minor importance because only the very small lateral motion [see Eq. (3)] that is normal to the surface generates sound.

For bars, rods, or columns of cross section S, density ρ, and Young's modulus E the equation of motion for the (longitudinal) displacement ξ_l in the direction of wave propagation (see Fig. 1a) is

---

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker
Copyright © 1997 John Wiley & Sons, Inc.

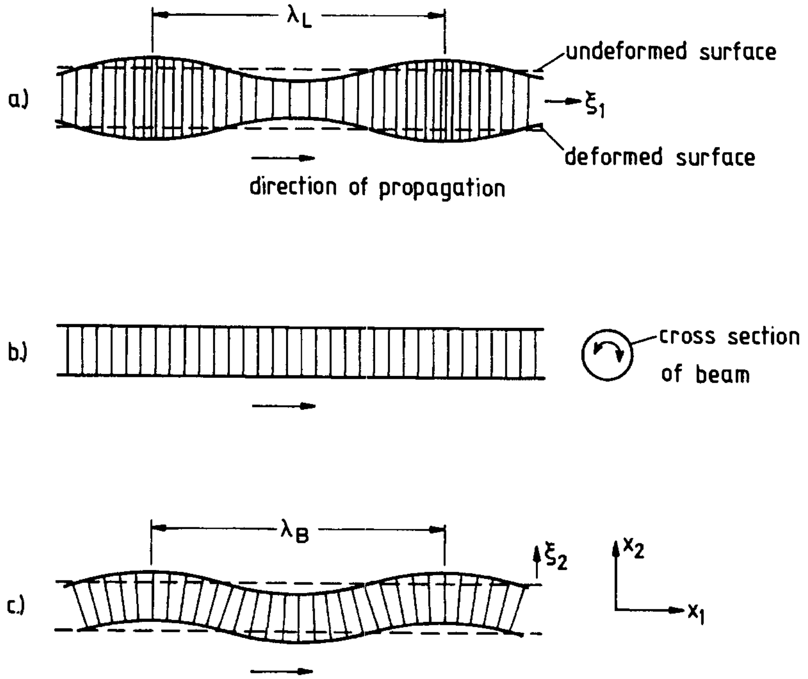


Fig. 1 Different wave types: (*a*) longitudinal with lateral contraction; (*b*) torsional; (*c*) bending.

$$ES\frac{\partial^2\xi_1}{\partial x_1^2} - \rho S\frac{\partial^2\xi_1}{\partial t^2} = 0\ [= -p_1(x_1)]. \qquad (1)$$

The term $p_1(x_1)$ represents forces per unit length that may be acting on the bar. (A rather unlikely case; therefore it is in brackets). As usual, t represents time.

Equivalent to Eq. (1) and sometimes more useful are the equations

$$F = -ES\frac{\partial\xi_1}{\partial x_1}, \qquad -\rho S\frac{\partial^2\xi_1}{\partial t^2}\,[+p_1(x_1)] = \frac{\partial F}{\partial x_1}. \qquad (2)$$

Here F is the force acting at x_1 over the cross section S.

The motion ξ_2 in the x_2-direction, which is due to lateral contraction in the case of harmonic motion, is given at the surface by

$$|\xi_2| = \frac{|\xi_1|\mu\pi h}{\lambda_L}. \qquad (3)$$

(A similar relation holds for the x_3-direction.) Poisson's ratio μ relates Young's modulus E and shear modulus G via the equation

$$G = \frac{E}{2 + 2\mu}. \qquad (4)$$

Here, h is the thickness and λ_L the longitudinal wavelength [see Eq. (8a)]. Some values for E and G are given in Table 1 in Chapter 62. The above equations are valid as long as $\lambda_L > 10h$. In the two-dimensional case lateral excitation leads not only to longitudinal but also to shear motion. This case is treated in Section A.3.

Bending Waves in Plates* For most practical applications bending waves in plates or platelike structures are the most important type of motion of two-dimensional structures from the acoustic point of view. Because of their comparatively low input impedance, they can be excited quite easily. In addition, they are good sound radiators (if their wavelength is not too short, see Chapter 11) because their motion is normal to the plate surface. Bending waves are very important for sound isolation problems in connection with walls, ceilings, enclosures, and so on.

For homogeneous plates in bending motion the displacement ξ_2 in the direction normal to the surface is given by[1–4]

$$D\left[\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_3^2}\right]^2 \xi_2 + \rho h\frac{\partial^2\xi_2}{\partial t^2}$$

$$= D\Delta\Delta\xi_1 + \rho h\frac{\partial^2\xi_2}{\partial t^2} = p(x_1, x_3), \qquad (5)$$

---

*For bending waves in beams see Section A.2 in the Appendix.

where $D = Eh^3/[12(1-\mu^2)]$ is the bending stiffness of the plate of thickness h, Δ is the Laplace operator, and $p(x_1, x_3)$ represents the external driving pressure (pressure difference between the two sides of the plate). The equation is valid in the range $\lambda_B > 6h$, where λ_B is the bending wavelength [see Eq. (9)]. For localized excitation one also has to demand that $p(x_1, x_3)$ does not change too rapidly in space because the effects of local elasticity are not included in Eq. (5). So-called point excitation should therefore be distributed over an area comparable to h^2. Otherwise, good correlation with measurements cannot be expected. The range of the applicability of bending theory can be somewhat extended by including the effects of shear stiffness and rotary inertia, as is done in Section A.4. In Section A.5 the related problem of orthotropic plates is also treated.

2.2 Dispersion Relations and Wave Speeds

From here on longitudinal waves in a bar (index L) and bending waves in a plate (index B) are treated simultaneously. They are considered to be the most important examples of one- and two-dimensional systems. For other examples some relevant equations are given in the Appendix. If there is no external excitation (i.e., the right-hand side of the equation of motion is zero) or if the excitation is outside the region of interest, there is still a nontrivial solution to the wave equations. It is called the free-wave solution. It can exist only when there is a certain relation (the dispersion relation) between frequency and wavenumber or wavelength or wave speed.

It is known that a linear superposition of exponential functions is the proper solution of linear differential equations in rectilinear coordinates. In one dimension such a function is

$$\mathrm{Re}\{\xi_1^v e^{j\omega t} e^{-jk_1 x_1}\}. \tag{6a}$$

In two dimensions one has

$$\mathrm{Re}\{\xi_2^v e^{j\omega t} e^{-jk_1 x_1} e^{-jk_3 x_3}\}. \tag{6b}$$

In the following the symbol Re $\{\cdot\}$ is not written explicitly and quite often the factor $\exp(j\omega t)$ is omitted. The terms ξ_1^v and ξ_2^v are complex amplitudes. To distinguish them from the time- and space-dependent quantities, they have the superscript v.

Equations (6a) and (6b) constitute a very convenient notation, especially when damping is involved. Care has to be taken, however, when energies, intensities, or similar quantities are calculated. In physical terms Eqs. (6a) and (6b) represent plane, harmonic waves with angular frequency ω and wavenumbers k_1, k_3. If Eqs. (6a) and (6b) are inserted into Eq. (1) or (5) with the right-hand side vanishing, the results are

$$Ek_{1f}^2 = \omega^2 \rho,$$

$$k_{1f}^2 = \frac{\omega^2 \rho}{E} = \frac{\omega^2}{c_l^2} \quad \text{(bar, longitudinal)}, \tag{7a}$$

$$D(k_{1f}^2 + k_{3f}^2)^2 = \omega^2 \rho h,$$

$$(k_{1f}^2 + k_{3f}^2)^2 = \frac{\omega^2}{c_L^2}\,\frac{12(1-\mu^2)}{h^2} \quad \text{(plate, bending)}. \tag{7b}$$

These are the so-called dispersion relations, which give the free wavenumbers k_{1f}, k_{3f} for a given frequency. Free wavenumbers are characterized by the fact that they can exist even for vanishing excitation, or equivalently that the wave impedance Z_T (see Section 2.5) becomes zero. Dispersion relations for other structures are given in the Appendix. In one dimension there is only one free wavenumber. It is related to the free wavelength λ_L and the phase speed c_{Ph} of propagation by

$$\lambda_L = 2\pi/k_{1f} = \frac{c_l}{f}, \qquad c_{\mathrm{Ph}} = \frac{\omega}{k_{1f}} = c_L. \tag{8a}$$

In two dimensions the situation is a little more complicated. For the plate the dispersion relations following from Eq. (7b) are

$$k_{1f}^2 + k_{3f}^2 = \frac{\pm\omega}{c_L h}\sqrt{12(1-\mu^2)} = \pm k_B^2. \tag{8b}$$

When k_{1f}^2 and k_{3f}^2 are sufficiently small, one can write them as $k_{1f} = k_B \cos\vartheta$ and $k_{3f} = k_B \sin\vartheta$. Thus Eq. (8b) shows that, when the plus sign is taken, there are free waves with a wavenumber k_B that can travel in any direction ϑ. When either k_{1f}^2 or k_{3f}^2 is larger than k_B^2 or when the minus sign is taken, either k_{1f} or k_{3f} or both are imaginary. In this case there is a so-called near-field or evanescent wave that decays exponentially.

Near fields are only important in the immediate vicinity of the excitation region or near boundaries because they decay by approximately 55 dB within one wavelength.

Wavelength λ_B and wave speed c_B in the direction of propagation are given by

$$\lambda_B = \frac{2\pi}{k_B} = 2\pi\sqrt{\frac{c_1^2 h^2}{[12\omega^2(1-\mu^2)]}}, \tag{9}$$

$$c_{\mathrm{Ph}} = c_B = \frac{\omega}{k_B} = \sqrt[4]{\frac{\omega^2 h^2 c_L^2}{12(1-\mu^2)}} = \sqrt[4]{\frac{\omega^2 h^2 E}{12\rho(1-\mu^2)}}. \tag{10}$$

Since the phase speed for bending waves depends on frequency, the group velocity c_{Bg}, which is the speed of propagation of the peak in a wave packet, is different. It is given by

$$c_g = \frac{d\omega}{dk_f} = \left(\frac{dk_f}{d\omega}\right)^{-1}. \tag{11}$$

For bending waves this yields

$$c_{Bg} = 2c_B.$$

2.3 Damping and Decay Rates

The simplest way to include material damping in the calculations is to introduce the loss factor η and to use complex moduli, complex wavenumbers, and complex amplitudes. In doing this, one is restricted to harmonic motion of frequency ω. If the complex Young's modulus

$$\underline{E} = E(1 + j\eta) \tag{12}$$

is introduced into Eqs. (1) and (5), the only change is that all quantities become complex. In particular, one finds

$$k_L^2 = \frac{\omega}{c_L} = \frac{\omega^2\rho}{E(1+j\eta)},$$

$$\underline{k}_L = k_L\left(1 - \frac{j\eta}{2}\right) \quad \text{(longitudinal)} \tag{13a}$$

$$k_B^2 = \sqrt{\frac{12\omega^2\rho(1-\mu^2)}{Eh^2}},$$

$$k_B = k_B\left(1 - \frac{j\eta}{4}\right) \quad \text{(bending).} \tag{13b}$$

The approximations for k_L and k_B hold if the damping is small; that is, $\eta^2 \ll 1$. (Here, underlined quantities are complex; quantities that are not underlined are real). Combination of Eqs. (13a) and (13b) with Eqs. (1) and (6) leads to decaying waves with amplitude dependence: For longitudinal waves

$$|\xi_1| \sim e^{-\eta k_1 x_1/2}; \tag{14a}$$

for bending with $\vartheta = 0$, that is, for plane waves in the direction of propagation,

$$|\xi_2| \sim e^{-\eta k_n x_1/4}. \tag{14b}$$

On a decibel scale this corresponds to a level difference of ΔL for a separation Δx:

$$\Delta L = \begin{cases} 27.28\eta\,\dfrac{\Delta x}{\lambda_L} & \text{(longitudinal)} \qquad (15a) \\[2ex] 13.64\eta\,\dfrac{\Delta x}{\lambda_B} & \text{(bending)} \qquad (15b) \end{cases}$$

The formulas hold when there is propagation in just one direction without any reflections or other disturbances. In practice, this is achieved only on large, damped structures. When point loads on plates generate circular waves, there is an additional decay ΔL_g due to geometric spreading. When R_1 and R_2 are the distances to the center of a localized excitation, it is given by

$$\Delta L_g = 10\log_{10}\left(\frac{R_2}{R_1}\right) \quad \text{(dB).} \tag{15c}$$

The exponential decay of energy on a very large structure with losses is just one consequence of damping. Other effects are the additional phase shift between force (stress) and displacement, the reduction of the mean-square velocity of a finite structure (see statistical energy analysis in Chapters 76 and 78), the decrease of reverberation time and resonance amplitude, and so on.

2.4 Energy

The energy density W_e consists of kinetic energy $W_{e,\mathrm{kin}}$ and potential energy $W_{e,\mathrm{pot}}$. For the two cases considered here they are:

$$W_e = W_{e,\mathrm{kin}} + W_{e,\mathrm{pot}} = \frac{1}{2}\rho S\xi_1^2 + \frac{1}{2}ES\left(\frac{\partial\xi_1}{\partial x_1}\right) \quad \text{(longitudinal),} \tag{16a}$$

$$\begin{aligned} W_e &= W_{e,\mathrm{kin}} + W_{e,\mathrm{pot}} \\ &= \frac{1}{2}\rho h \xi_2^2 + \frac{1}{2} D\left\{\left(\frac{\partial^2 \xi_2}{\partial x_1^2} + \frac{\partial^2 \xi_2}{\partial x_3^2}\right)^2\right. \\ &\quad \left. - 2(1-\mu)\left[\frac{\partial^2 \xi_2}{\partial x_1^2}\,\frac{\partial^2 \xi_2}{\partial x_3^2} - \left(\frac{\partial^2 \xi_2}{\partial x_1\,\partial x_3}\right)^2\right]\right\} \\ &\quad \text{(bending).} \end{aligned} \tag{16b}$$

Here $\dot{\xi}_1$ and $\dot{\xi}_2$ are the particle velocities, that is, the time derivatives of ξ_1 and ξ_2. Note that Eq. (16a) has the dimension energy per length (watts by seconds per metre), and Eq. (16b) has energy per area (watts by seconds per metre square).

If in Eqs. (16a) and (16b) the potential energy is not added but subtracted, the Lagrangian

$$L = W_{e,\mathrm{kin}} - W_{e,\mathrm{pot}} \tag{16c}$$

is obtained. This is a very useful quantity when it is applied in connection with Hamilton's principle. See Chapter 62.

Another closely related quantity, the intensity, which determines the energy flow in structures, is not treated here because it is the subject of Chapter 74.

2.5 Wave Impedance for Excitation by Plane Waves

Since one- and two-dimensional structures can be excited in many different ways, it seems appropriate to consider only a few idealized cases. One of them is the excitation of an infinite homogeneous structure by a plane wave with given frequency and wavenumber (or wave speed). The advantage of this approach is that it can be generalized, because all types of plane excitation can be expressed as a sum of plane waves; that is, the Fourier transform of the excitation can be taken. There remains, however, the restriction that the excited structure has to be very large and homogeneous. The general idea underlying the procedure is the following[2,3]:

(a) Because of linearity, all quantities are expressed as a sum of plane progressive waves, that is, as Fourier transforms.

(b) If the response of the structure to one plane wave is known (this is the purpose of the so-called wave impedance), the total response is simply an integration of many such responses.

If one applies Fourier transforms, the external pressure and the displacement can be written as

$$\begin{aligned} p_1(x_1) &= \frac{1}{2\pi}\int p_1^v(k_1)e^{-jk_1x_1}\,dk_1, \\ \xi_1(x_1) &= \frac{1}{2\pi}\int \xi_1^v(k_1)e^{-jk_1x_1}\,dk_1 \end{aligned} \tag{17a}$$

for one dimension and

$$\begin{aligned} p(x_1,x_3) &= \frac{1}{4\pi^2}\int p^v(k_1,k_3)e^{-jk_1x_1}e^{-jk_3x_3}\,dk_1\,dk_3, \\ \xi_2(x_1,x_3) &= \frac{1}{4\pi^2}\int \xi_2^v(k_1,k_3)e^{-jk_1x_1}e^{-jk_3x_3}\,dk_1\,dk_3. \end{aligned} \tag{17b}$$

for two dimension. (All integrations are in the range $-\infty$ to $+\infty$.)

If these expressions are inserted in Eqs. (1) and (5), the result is

$$\begin{aligned} \xi_1^v(k_1) &= \frac{p_1^v(k_1)}{ESk_1^2 - \omega^2\rho S} = \frac{p_1^v(k_1)}{j\omega}\,\frac{1}{j\omega\rho S - jESk_1^2/\omega} \\ &= \frac{p_1^v(k_1)}{j\omega}\,\frac{1}{Z_{\tau L}} \end{aligned} \tag{18a}$$

for longitudinal motion of a bar and

$$\begin{aligned} \xi_2^v(k_1,k_3) &= \frac{p^v(k_1,k_3)}{D(k_1^2+k_3^2)^2 - \omega^2\rho h} \\ &= \frac{p^v(k_1,k_3)}{j\omega}\,\frac{1}{j\omega\rho h - j(k_1^2+k_3^2)^2 D/\omega} \\ &= \frac{p^v(k_1,k_3)}{j\omega}\,\frac{1}{Z_{\tau B}} \end{aligned} \tag{18b}$$

for bending motion of a plate. The quantities

$$Z_{\tau L} = j\left(\omega\rho S - \frac{ESk_1^2}{\omega}\right) \quad \text{(for bars)}, \tag{19a}$$

$$Z_{\tau B} = j\left[\omega\rho h - \frac{(k_1^2+k_3^2)^2 D}{\omega}\right] \quad \text{(for plates)} \tag{19b}$$

are the wave impedances. Sometimes they are called transmission impedances. They relate the amplitude of an exciting plane pressure wave to the amplitude of the plane velocity wave that is generated this way in the structure.

The equations show that there are certain wavenumbers for which the wave impedance becomes zero; that is, a very small exciting force is sufficient to cause large vibrations. This phenomenon is called coincidence.[6] It occurs when the wavenumber of the excitation coincides with a free wavenumber of the structure or when the excitation moves with a speed that is very close to a free wave speed in the structure. Coincidence is of practical importance for the sound isolation of walls because sound waves that have a trace wave speed equal to the bending wave speed in a wall are transmitted almost unattenuated. Coincidence phenomena can also be important for the excitation of structures by turbulent boundary layers when the flow speed equals a free wave speed. Another example of coincidence can occur when the rolling speed (e.g., of a tire) is equal to a free wave speed in a structure that is in contact.

The existence of coincidence is shown here only for infinite structures. In finite structures it is also found but is less pronounced, especially when the dimensions of the structure are less than a wavelength.

2.6 Sound Transmission Loss

If the procedure of the last section is applied to the sound transmission problem, the driving pressure $p(x_1, x_3)$ consists of the incoming wave p_i plus the reflected wave p_r minus the transmitted wave p_t.[5,6] With the symbols used in Fig. 2 this gives the following expressions for the transmission coefficient τ and the sound reduction index R (also known as transmission loss TL):

$$\tau = \left|\frac{p_t}{p_i}\right|^2 = \left|1 + \frac{Z_\tau}{2\omega\rho_0}\sqrt{k_0^2 - k_1^2 - k_3^2}\right|^{-2},$$

$$R = 10\log_{10}\frac{1}{\tau} \qquad \text{(dB)}. \tag{20}$$

Here Z_τ is the wave impedance of the structure. Complete transmission with $\tau = 1$ and $R = 0$ occurs at coincidence; that is, when the free wavelength of the structure coincides with the trace wavelength $\lambda_0/\sin\vartheta$ of the incoming sound. See also Chapter 11.

In practical situations, when sound is coming from all directions to a wall, Eq. (20) has to be averaged over all angles of incidence. This has to be done with care, because τ changes extremely rapidly as the angle of incidence changes near coincidence. For a large, plane, homogeneous plate in bending the average result is

$$R \approx \begin{cases} 20\log_{10}\dfrac{\omega\rho h}{2\rho_0 c_0} - 5 \\ \quad \text{if} \quad \omega\rho h \gg \rho_0 c_0, \quad \omega < \omega_g, \\ 20\log_{10}\dfrac{\omega\rho h}{2\rho_0 c_0} + 10\log_{10}\dfrac{\omega\eta}{\omega_g} - 3 \\ \quad \text{if} \quad \omega\rho h \gg \rho_0 c_0, \quad \omega > 2\omega_g. \end{cases} \tag{21}$$

Here

$$\omega_g = c_0^2\sqrt{\frac{\rho h}{D}} \tag{22}$$

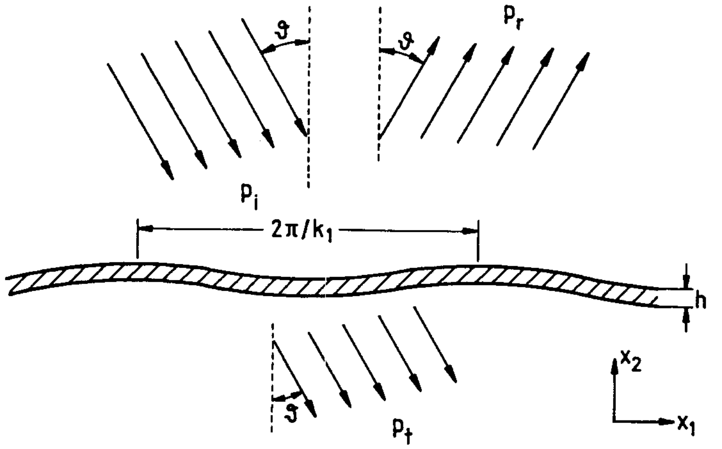


Fig. 2 Sound transmission through a plane structure of infinite extent: $k_0 = \omega/c_0$; $\cos\vartheta = \sqrt{k_0^2 - k_1^2 - k_3^2}/k_0$; ρ_0 and c_0 are density and speed of sound of the surrounding gas or fluid.

is the limiting frequency of coincidence where bending wave speed and speed of sound in the surrounding medium are equal.

The formulas agree fairly well with measured data, even for finite walls with dimensions of several wavelengths, provided the boundary losses are somehow included in the loss factor η. More on sound isolation is presented in Chapter 93.

Equation (20) does not include the effect of shear stiffness and rotary inertia. If this were done, it would result in slightly smaller values of R at high frequencies.

2.7 Point Impedances

In Section 2.5 the general method for calculating the motion of very large plane structures was described. Since it involves at least one integration, it is useful to give results for special cases. One such special case is point excitation with a localized force of frequency ω. Localized (i.e., "point") excitation in this context means that the excited area is small compared with the wavelength but not small compared with the thickness.

Results obtained in this way for excitation with a harmonic force of amplitude F_0 are

$$Z_{FL} = \frac{F_0}{v_0} = \epsilon \rho c_L S \quad \text{(bar in longitudinal motion)}, \tag{23a}$$

$$Z_{FB} = \frac{F_0}{v_0} = 8\sqrt{\rho h D} = \frac{8\omega\rho h}{k_B^2} \quad \text{(plate in bending)}. \tag{23b}$$

Here Z_F is the point impedance, which is defined as the (complex) ratio of the force amplitude to the velocity amplitude v_0 at the excitation point. $\epsilon = 1$ for a semi-infinite bar and $\epsilon = 2$ for an infinite bar. For a semi-infinite plate excited at the edge the factor 8 in Eq. (23b) has to be replaced by 2.35.

The point impedances for beams in bending and for several other structures are given in the Appendix.

The mechanical power that is transmitted by a force is [see also Eq. (14)]

$$\begin{aligned} P_F &= \overline{\mathrm{Re}\,\{F_0 e^{j\omega t}\}\,\mathrm{Re}\{v_0 e^{j\omega t}\}} = \frac{1}{2}\,\mathrm{Re}\{F_0 v_0^*\} \\ &= \frac{1}{2}\,|F_0^2|\,\mathrm{Re}\left\{\frac{1}{Z_F}\right\}. \end{aligned} \tag{24}$$

The asterisk indicates complex-conjugate values. The overbar indicates time averaging.

When a homogeneous plate is excited by two opposite forces that are a distance apart, there exists "moment excitation." In this case a moment impedance Z_M can be defined by

$$\begin{aligned} Z_M &= \frac{M_0}{w_0} = \frac{2Z_F}{k_B^2}\left[1 + j\,\frac{4}{\pi}\,\ln\frac{1.1}{k_B a}\right], \\ P_M &= \frac{1}{2}\,|M_0^2|\,\mathrm{Re}\left\{\frac{1}{Z_M}\right\} \approx \frac{1}{8}\,|M_0^2|\,\mathrm{Re}\left\{\frac{1}{Z_F}\right\}k_B^2. \end{aligned} \tag{25}$$

Here M_0 is the exciting moment and $w_0 = j\omega\,\partial\xi_2/\partial x_1$, the amplitude of the angular velocity at the excitation point in the direction of the force separation.

The point impedances given here hold for infinite structures. But is can be shown that the quantity $\mathrm{Re}\{1/Z_F\}$, which is very important for power transmission, is also applicable for finite structures, provided that frequency averages are taken that contain at least two resonance frequencies. For practical purposes it may be useful to note that impedances become bigger when the mass and/or the stiffness is increased.

For engineering applications the *power transfer quantity* $\mathrm{Re}\{1/Z_F\}$ can be estimated in a rather crude way by

$$\mathrm{Re}\left\{\frac{1}{Z_F}\right\} \approx \frac{1}{\omega\rho h S_q}. \tag{26}$$

Here ρ is the density of the material and S_q the size of the area that moves with the excitation. It is given by

$$S_g \approx \begin{cases} b\lambda_1/3 & \text{for one-dimensional structures,} \\ (\lambda_1/3)(\lambda_3/3) & \text{for two-dimensional structures.} \end{cases}$$

Here b is the width of the structure; λ_1, λ_3 are the wavelengths of the relevant structural waves in the two directions. It also can be shown that the real part of the inverse moment impedance (i.e., the moment mobility) is usually proportional to the real part of the inverse force impedance (i.e., the force mobility) divided by the square of the wavelength. Thus moment mobilities tend to become bigger with frequency. Consequently the relative importance of moment excitation as compared to force excitation increases with frequency.

2.8 Green's Functions

With the methods described in Section 2.5 it is also possible to find the motion of a structure when it is excited by a localized source. The function obtained this way is the Green's function. For a semi-infinite bar in longitudinal motion excited at one end by a harmonic force, the result is

$$v(x_1) = j\omega\xi_1(x_1) = F_0\rho c_L S\, e^{-jk_L x}. \tag{27a}$$

If the excitation is due to a sudden pulse of type $F_0(t) = I\delta(t)$, the response is

$$v(x_1, t) = \dot{\xi}_1(x_1, t) = I\rho c_L S\delta\left(t - \frac{x_1}{c_1}\right). \tag{27b}$$

This shows that the pulse I travels unchanged with speed c_L through the bar until it hits some sort of discontinuity or boundary where it gets fully or partially reflected.

The corresponding results for an infinite plate in bending are

$$v(r) = j\omega\xi_2(r) = \frac{F_0[H_0^{(2)}(k_B r) - H_0^{(2)}(-jk_B r)]}{Z_F}$$

$$\text{(harmonic excitation)}, \tag{27c}$$

$$v(r, t) = j\omega\xi_2(r, t) = \frac{I}{4\pi\sqrt{D\rho h}\,t}\sin\left(\frac{r^2}{4t}\sqrt{\frac{\rho h}{D}}\right)$$

$$\text{(pulse excitation)}. \tag{27d}$$

Here $\delta(\cdot)$ is the Dirac delta function, $H_0^{(2)}(\cdot)$ the Hankel function, I the transmitted momentum, and r the distance between excitation and observation.

Equation (27c) shows that for a given frequency bending waves on a very large plate decay with the square root of the distance, as is required by energy considerations. Because of the dependence on $k_B r = 2\pi r/\lambda_B$, the decay is more pronounced at high frequencies.

Equation (27d) shows that a short pulse gets completely distorted as it travels away from the source point. After some travel time it consists of small, rather rapid (high-frequency) fluctuations followed by larger and slower vibrations.

Note that in the harmonic case [i.e., Eqs. (27a) and (27c)] the time dependence $\exp(j\omega t)$ has to be taken into account and the real part has to be taken if the time history of the velocity is the desired quantity.

Due to the limits of applicability of the equations of motion, the results hold only below a certain frequency limit and not for very short times. For $k_B r > 2$, Eq. (27c) becomes

$$v(r) \approx \frac{F_0}{Z_F}\sqrt{\frac{2}{\pi k_B r}}\, e^{-jk_B r} e^{j\pi/4}. \tag{27e}$$

This resembles the well-known Green's function for sound waves in a two-dimensional free space.

2.9 Normal Modes and Resonance Frequencies

There are several possibilities to treat the influence of boundaries on the propagation of elastic waves. One method is to follow the paths of the waves including all reflections. This is equivalent to adding up the waves that are generated by the original source and all its images (see Chapter 91). This way the problem is reduced to investigating an infinite structure with many sources. Resonances occur when the waves originating from the images add up in phase. If the integral in Eq. (17b) is replaced by a Fourier series, that is, if there is a summation symbol instead of the integral, $(2\pi)^2/l_1 l_3$ instead of $dk_1\, dk_3$, and $2\pi n_1/l_1$, $2\pi n_3/l_3$ instead of k_1, $k_3 (n_1, n_3 = 1, 2, 3, \ldots)$, one is in essence dealing with a configuration that repeats itself with periodicity l_1 and l_3. Thus one has exactly the same situation that is obtained by image formation. The problem, however, is that one is restricted to rectangular structures and there is no way to include boundary conditions other than perfect reflection.

A second method is to express the wave field on a finite, homogeneous structure as a sum of possible propagating waves and near fields in the position and negative directions. For a beam in bending this would lead to

$$\xi(x) = A_+ e^{-jk_f x} + A_- e^{jk_f(x-1)} + B_+ e^{-k_N x} + B_- e^{k_N(x-1)}. \tag{28}$$

The two ends of the beam are at $x = 0$ and $x = 1$. Here, k_f is the free wavenumber for propagating waves and k_N is the near-field wavenumber. When the simple Bernoulli–Euler theory of bending is used (see Section A.2), the wavenumbers are given by $k_f = k_N = k_B$. When the more-accurate Timoshenko theory is used, one has to use the equations in Section A.4 of the Appendix and set $k_f = k_I$ and $jk_N = k_H$. The terms A_+, A_-, B_+, B_- are unknown coefficients that depend on the boundary conditions.

The corresponding method for plates consists in expressing the velocity field as a sum of propagating cylindrical waves with wavenumber k_F and cylindrical near fields with wavenumber jk_N. The amplitudes of these waves and near fields is then adjusted in such a way that a best fit to the boundary conditions is obtained.

A third method is the analog to the "solution" of the sound wave equation by the Kirchhoff–Helmholtz integral equation. In this method the wave field inside a boundary is expressed by a surface integral over the driving (outside) pressure $p(x_1, x_3)$ and a line integral over the field quantities along the boundary.[7]

The fourth and at present probably most often used method[8] to deal with the vibrations of one- and two-dimensional continuous systems is the finite element method (FEM) (see Chapters 14 and 63). This method, which can be seen as an application of Hamilton's principle, is described elsewhere.

For analytical calculations it is sometimes useful[9–11] to have simple expressions for the normal modes of a structure. For one-dimensional beams it is always possible to find the normal modes by using Eq. (28) and looking for solutions that fulfil the boundary conditions and that can exist without external excitation. For rods or bars exhibiting longitudinal or torsional motion, the results are the same as those for sound waves in tubes (see Chapter 7); the only difference is that the speed of sound has to be replaced by the longitudinal or torsional wave speed. For specific results on resonance frequencies and mode shapes the reader is referred to Refs. 1–3 and 9–11.

3 THIN, CURVED STRUCTURES

3.1 Cylinders

For the thin cylindrical shell shown in Fig. 3 there exist in the literature many slightly different equations of motion, depending on the required degree of accuraccy.[1–11] A simple and for most cases sufficiently accurate equation is

$$\begin{aligned}
&a_{11}u + a_{12}v + a_{13}w = 0,\\
&a_{12}u + a_{22}v + a_{23}w = 0,\\
&a_{13}u + a_{23}v + a_{33}w = \frac{1-\mu^2}{Eh}\,p_r,
\end{aligned}$$

$$\begin{aligned}
a_{11} &= \frac{\partial^2}{\partial z^2} + \frac{1-\mu}{2a^2}\frac{\partial^2}{\partial \vartheta^2} - \frac{1}{c_L^2}\frac{\partial^2}{\partial t^2},\\
a_{12} &= \frac{1+\mu}{2a}\frac{\partial^2}{\partial z\,\partial \vartheta},\\
a_{13} &= \frac{\mu}{a}\frac{\partial}{\partial z},\\
a_{22} &= \frac{1-\mu}{2}(1+\beta^2)\frac{\partial^2}{\partial z^2} + \frac{1}{a^2}(1+\beta^2)\frac{\partial^2}{\partial \vartheta^2}\\
&\quad - \frac{1}{c_L^2}\frac{\partial^2}{\partial t^2},\\
a_{23} &= \frac{1}{a^2}\frac{\partial}{\partial \vartheta} - \beta^2\left(\frac{\partial^3}{\partial z^2\,\partial \vartheta} + \frac{1}{a^2}\frac{\partial^3}{\partial \vartheta^3}\right),\\
a_{33} &= \frac{1}{a^2} + \frac{h^2}{12a^2}\left(a^2\frac{\partial^4}{\partial z^4} + 2\frac{\partial^4}{\partial z^2\,\partial \vartheta^2} + \frac{1}{a^2}\frac{\partial^4}{\partial \vartheta^4}\right)\\
&\quad + \frac{1}{c_L^2}\frac{\partial^2}{\partial t^2},
\end{aligned}\tag{29}$$

where u, v, w are the displacements in the axial, tangential, and radial directions; p_r is the pressure that excites the cylinder; and c_L is given by $c_L^2 = E/\rho(1-\mu^2)$.

For many applications one can neglect terms with $\beta^2 = h^2/12a^2$. In this way the Donell–Mushtari shell equation is obtained. It is the simplest shell equation that, in addition to the membrane stresses, takes into account bending stresses in an approximate way.

The free wavenumbers (dispersion relation) of Eq. (29) can be obtained by inserting

$$\begin{aligned}
u &= U\cos n\vartheta\, e^{-jk_z z}e^{j\omega t},\\
v &= V\sin n\vartheta\, e^{-jk_z z}e^{j\omega t},\\
w &= W\cos n\vartheta\, e^{-jk_z z}e^{j\omega t}
\end{aligned}\tag{30}$$

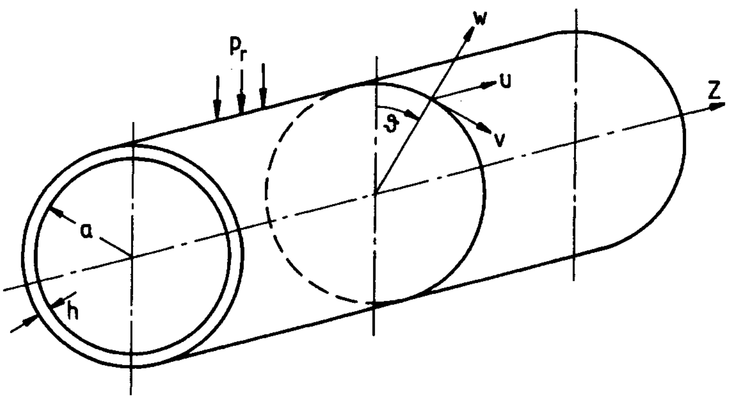


Fig. 3 Coordinate system for cylindrical shell.

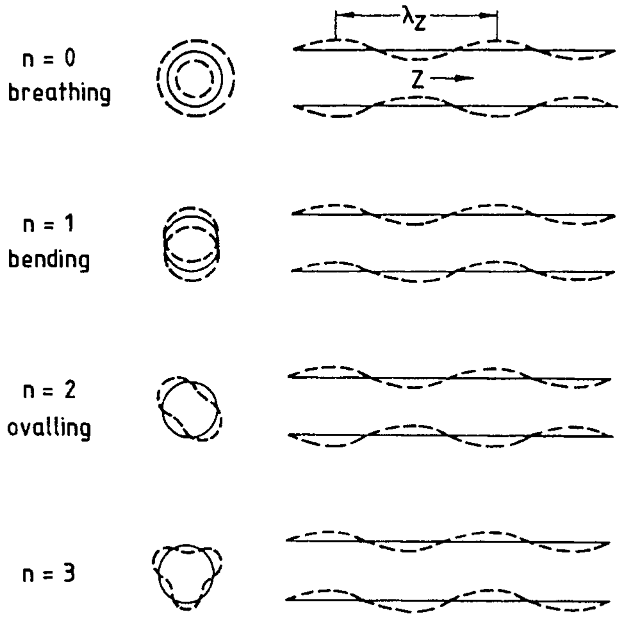


Fig. 4 Examples of mode shapes of a cylinder, $k_Z = 2\pi/\lambda_Z$.

into Eq. (29). In these equations $2n$ is the number of circumferential nodal lines and $2\pi/k_z$ the wavelength in the axial direction. See Fig. 4. One can view Eq. (30) also as two waves spinning in opposite directions around the cylinder. See also Section 4 in Chapter 11, where cylindrical shells are treated in connection with their sound radiation.

Combination of Eqs. (29) and (30) gives an expression of order $(k_z^2)^4$. Therefore there are four different waves or near fields. Closer examination shows that there are at most three different propagating waves for a particular value of n. For the case $a = 15h$ the wave speeds $c_z = \omega/k_z$ are plotted in Fig. 5. From this plot and several additional calculations the following can be seen:

(a) In the frequency range

$$\frac{\omega a}{c_L} < \frac{6\beta}{\sqrt{5}} \approx 0.775h/a$$

there are two waves with $n = 0$, one corresponding to a longitudinal motion with breathing, the other to a torsional motion. The only other possible (and very important) cylinder motion in this frequency range is bending ($n = 1$) as if the cylinder were a beam.

(b) In the range

$$\left(\frac{\omega a}{c_L}\right)^2 < \frac{1}{2}\,n^2(1 - \mu)$$

there is one wave for each $n \le 2$; it starts with an infinite phase velocity (and a zero group velocity) at the ring frequency

$$\frac{\omega_n a}{c_L} = \frac{h}{\sqrt{12}a}\,\frac{n(n^2 - 1)}{\sqrt{n^2 + 1}}. \tag{31}$$

These waves are closely related to bending waves of a plate of thickness h made out of the cylinder material.

(c) Starting from $(\omega a/c_L)^2 = \frac{1}{2}n^2(1 - \mu)$ there is a group of waves that corresponds to shear waves in a plate.

(d) Starting from $(\omega a/c_L)^2 = n^2 + 1$ there is a group of waves corresponding to longitudinal waves in a plate.

If in Eq. (29) the substitution $x = a\vartheta$ is made, it is seen that the equations decouple for $a \to \infty$ giving the wave equations for in-plane waves (see Section A.3) and for bending waves [see Eq. (5)]. Thus in a good approximation one can say that for $\omega a/c_L > 2$ the curvature effects are of minor importance and a cylinder behaves almost like a flat plate of thickness h.

3.2 Rings

The equation of motion for rings is obtained by setting all terms with $\partial/\partial z$ in Eq. (29) to zero. This way equa-

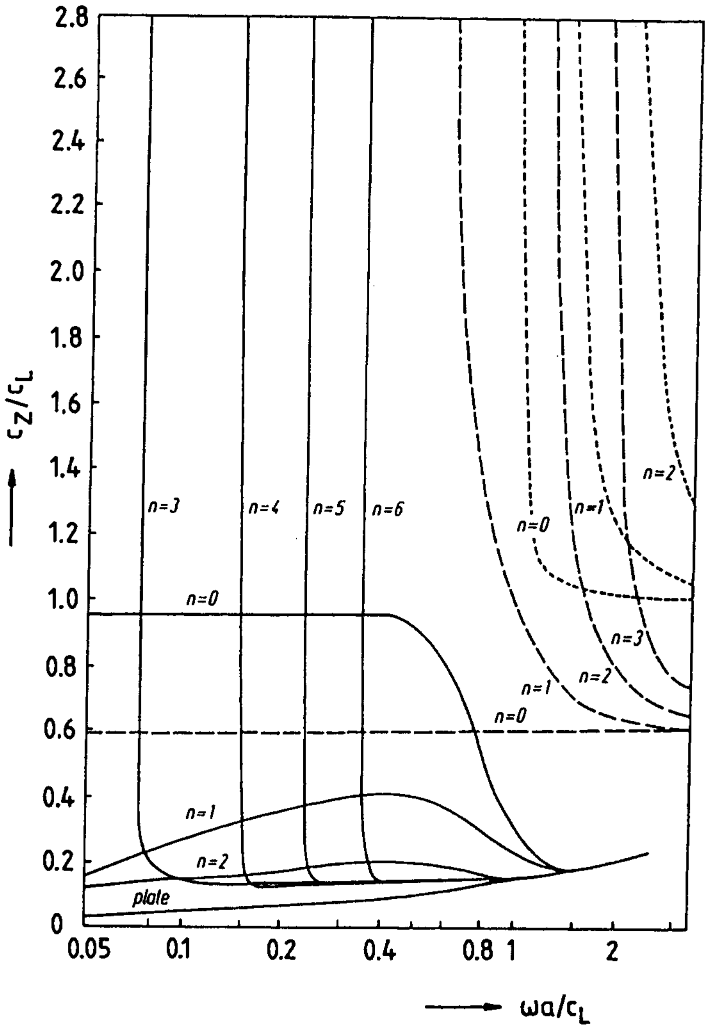


Fig. 5 Axial phase velocities of a cylinder with $a = 15h$; $\mu = 0.3$: The curve designated by "plate" corresponds to the phase velocity of flexural waves on a flat plate of thickness h. (-----) Related to shear motion. (-------) Related to longitudinal motion. (From Ref. 2.)

tions for the circumferential and radial motion (v, w) and for shear motion (u) in axial direction are obtained. The ring resonance frequencies that can be found from these expressions are

$$\omega_n = \frac{hc_L}{\sqrt{12a^2}} \frac{n(n^2 - 1)}{\sqrt{n^2 + 1}} \tag{32}$$

and (of much less interest because twisting is not included)

$$\omega_n = \frac{nc_r}{a}$$

4 COMBINATIONS OF SUBSTRUCTURES

4.1 Beams

Beams that are connected with each other can be considered as one-dimensional models of intersecting walls. For their mathematical treatment one can use the following method[3]:

(a) For each substructure the motion is expressed in terms of the free propagating waves and near fields, similar to Eq. (28).

(b) The boundary conditions at the intersection points and at the ends are used to calculate the unknown amplitudes.

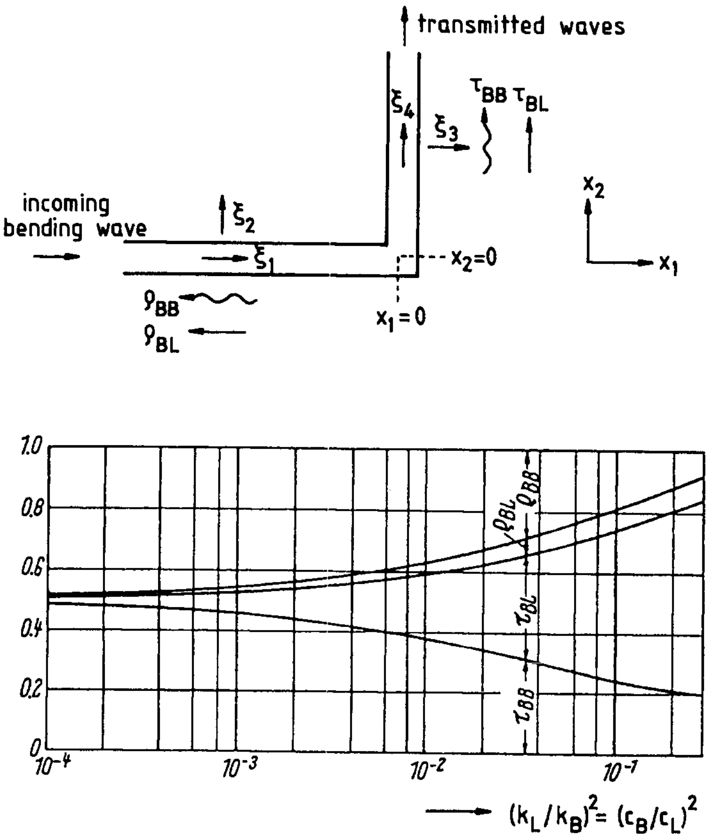


Fig. 6 Transmission of waves over a corner. The diagram was calculated for the case of identical beams.

(c) From these amplitudes the transmission coefficients can be found. For the example shown in Fig. 6 the following relations have to be used:

$$\xi_1(x_1) = r_{BL}e^{+jk_{L1}x_1},$$

$$\xi_2(x_1) = e^{-jk_{B1}x_1} + r_B e^{jk_{B1}x_1} + r_N e^{k_{B1}x_1} \qquad \text{(horizontal)},$$

$$\xi_4(x_2) = t_{BL}e^{-jk_{L2}x_2},$$

$$\xi_3(x_2) = t_B e^{-jk_{B2}x_2} + L_N e^{jk_{B2}x_2} \qquad \text{(vertical)}. \qquad (33)$$

The boundary conditions at $x_1 = 0$, $x_2 = 0$ are

$$\xi_1 = \xi_3,$$

$$\xi_2 = \xi_4 \qquad \text{(equality of displacements)},$$

$$\frac{\partial \xi_2}{\partial x_1} = \frac{\partial \xi_3}{\partial x_2} \qquad \text{(equality of angles)},$$

$$B_1 \frac{\partial^2 \xi_2}{\partial x_1^2} = B_2 \frac{\partial^2 \xi_3}{\partial x_2^2} \qquad \text{(equality of moments)},$$

$$-B_1 \frac{\partial^3 \xi_2}{\partial x_1^3} = -E_2 S_2 \frac{\partial \xi_4}{\partial x_2},$$

$$-E_1 S_1 \frac{\partial \xi_1}{\partial x_1} = -B_2 \frac{\partial^3 \xi_3}{\partial x_2^3} \qquad \text{(equality of forces)}.$$

These six conditions are sufficient to find t_B, r_B, ... The curves in Fig. 6 were obtained this way. They show the reflection and transmission coefficients for the reflected bending energy, the transmitted bending energy, the reflected longitudinal energy, and the transmitted longitudinal energy. All are normalized by the incoming energy:

$$\rho_{BB} = |r_B^2|, \qquad \tau_{BB} = \frac{\rho_2 S_2 k_{B1}}{\rho_1 S_1 k_{B2}} |t_B^2|,$$

$$\rho_{BL} = \frac{k_{B1}}{2k_{L1}} |r_{BL}^2|, \qquad \tau_{BL} = \frac{\rho_2 S_2 k_{B1}}{2\rho_1 S_1 k_{L2}} |t_{BL}^2|. \qquad (34)$$

It is interesting to note that, at least on a decibel scale, τ_{BB} is not very small. Thus corners are not much of an obstacle (typically 3 dB) for bending waves. At low frequencies there is almost no transition of bending wave energy to longitudinal waves, but at higher frequencies this effect can be substantial. An alternative method of treating beam combinations is to apply Hamilton's principle, which in essence corresponds to the substructure technique (with each beam being an element) in the FEM.

4.2 Multilayer Plates

Multilayer plates (see, e.g., Fig. 7) are very common as plywood, sandwich plates, or plates with damping layers. If it is assumed that the core is reasonably stiff so that for the frequencies of interest the motion $\xi_2(x_1, x_3)$ in the normal direction is the same over the whole thickness, such plates can be treated by standard techniques. A convenient method to find the equation of motion for multilayer plates is again Hamilton's principle. A typical result of such a calculation has the following general structure:

$$\alpha_{11}\xi_1 + \alpha_{12}\xi_2 = 0, \qquad \alpha_{12}\xi_1 + \alpha_{22}\xi_2 = p_2. \qquad (35)$$

Here the α_{ij} are rather lengthy linear differential operators that contain the different material properties and dimensions.

An interesting feature of Eq. (35) is the fact that in general the equations cannot be decoupled by introducing a frequency-independent neutral fiber for which $\xi_1 = 0$. Thus, in contrast to homogeneous beams, there is always a (small) coupling between bending motion and longitudinal motion.

By introducing a general plane-wave solution [see Eq. (6)], the dispersion relation, the free wavenumbers, and the wave speeds are obtained. By setting $\underline{G}_2 = G_2(1 + j\eta_{2G})$ and $\underline{E}_2 = E_2(1 + j\eta_{2E})$, where η_{2G} and η_{2E} are loss factors of the inner material, it is also possible to find the loss factor and decay rate of the multilayer beam. The results are practically the same as those given in Chapter 149, although there a quasistatic theory is used.

5 PERIODIC STRUCTURES

Periodic structures such as frame-stiffened ship hulls, airplane fuselages, certain types of ceiling constructions, and railway rails on ties are very common. For low frequencies, when the wavelength is much longer than the distance 1 between two discontinuities (see Fig. 8), periodic structures can be treated like homogeneous ones with modified mass (and rotary inertia). For discontinuities with a force impedance Z_{FD} on a bar, this can be done by replacing the mass per unit area ρS by the modified mass

$$\rho S\left(1 + \frac{Z_{FD}}{j\omega\rho S l}\right).$$

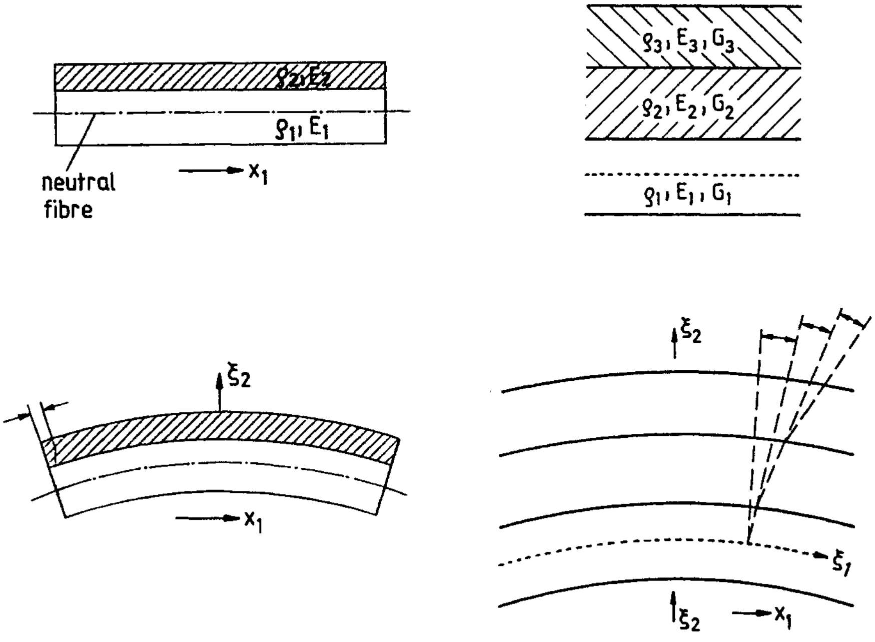


Fig. 7 Geometry of two- and three-layer plates.

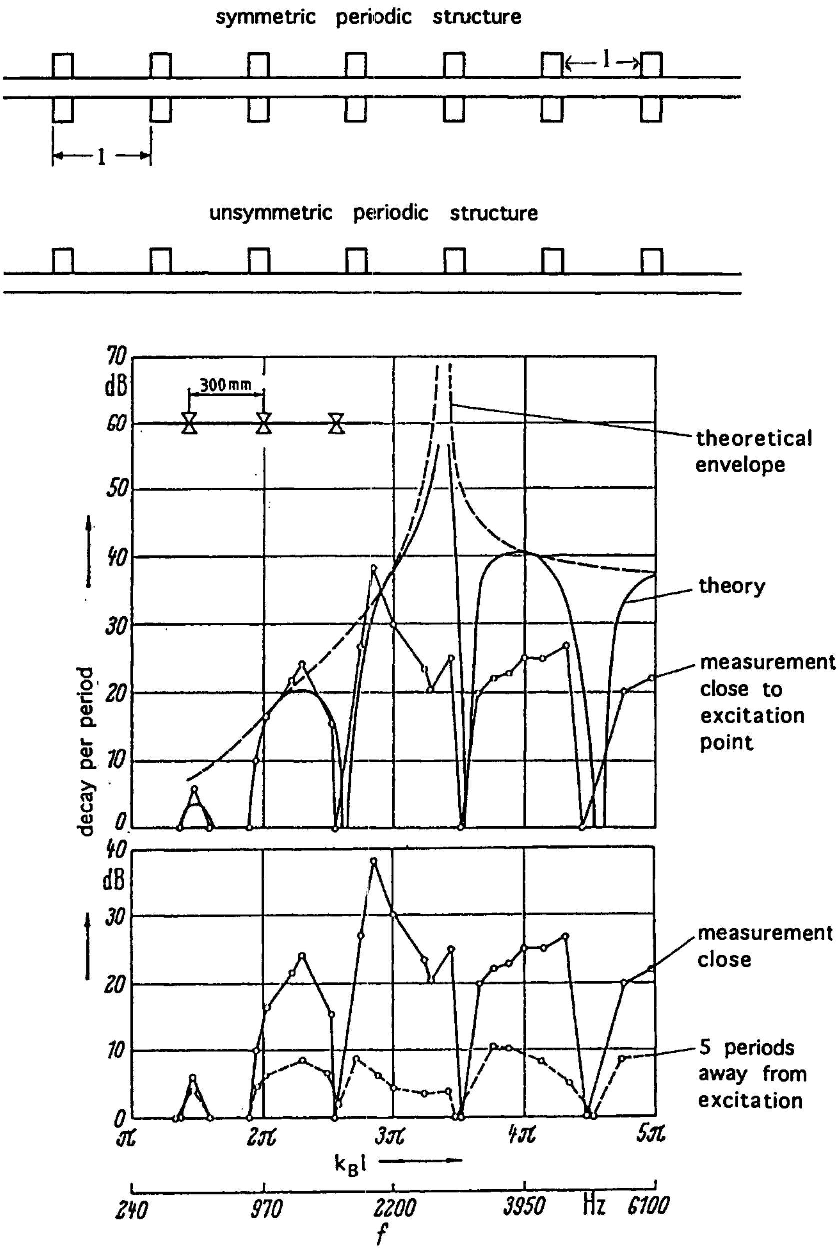


Fig. 8 Periodic system. The frequency curve holds for "blocking masses" of 600 g on a steel beam of cross section 10 × 30 mm.

For longitudinal waves on bars [see Eqs. (1) and (2)] this leads to the longitudinal, free wavenumber [instead of Eq. (7a)]

$$k_{lfD}^2 = \frac{\omega^2 \rho S}{ES}\left(1 + \frac{Z_{FD}}{j\omega\rho S l}\right) = k_{lf}^2\left(1 + \frac{Z_{FD}}{j\omega\rho S l}\right). \qquad (36)$$

This method works for sinusoidal motion of frequency ω, provided the discontinuities are almost symmetric (otherwise new wave types are excited). The impedance may consist of a single mass m_D (i.e., $Z_{FD} = j\omega m_D$) or a combination of mass, spring, and damper elements. It also may be the impedance of a continuous structure.

For beams in bending Eq. (36) is a poor approximation in certain frequency bands. It is better to include also the effect of rotary inertia. This leads to the following equation for the low-frequency wavenumber k_{BD} of a beam with symmetric periodic discontinuities of force impedance Z_{FD} and moment impedance Z_{MD}:

$$k_{BD}^2 \approx k_B^2 \left[\alpha \pm \sqrt{1 + \frac{Z_{FD}}{j\omega\rho S l} + \alpha^2} \right], \qquad (37)$$

with

$$\alpha = \frac{Z_{MD} k_B^2}{2j\omega\rho S l}.$$

Here k_B is the free bending wavenumber of the beam without the discontinuities. It is given in the Section A.2.

It can be easily seen that impedances of inertia type (i.e., proportional to $+j\omega$) decrease the wavelength; impedances of spring type (i.e., proportional to $-j/\omega$) increase it or may even prevent waves from propagating. Real-valued impedances cause damping.

When the frequency is high enough to make $|k_{BD} l| > 2$ or $|k_{LD} l| > 2$, Eqs. (36) and (37) are no longer valid. In this case one has to apply the well-developed theory of periodic systems.[2, 12, 13] Some interesting results obtained this way and verified by experiments are as follows:

(a) When the wavelength is comparable with or smaller than the period spacing, there are stopping and passing bands, that is, frequency ranges in which the waves decay exponentially in space and alternate frequency ranges in which they propagate almost without attenuation. An example is shown in Fig. 8.

(b) The actual location of the stopping and passing bands is a rather complicated function of the force and moment impedances of the added frames, ties, and so on.

(c) When the periodic structure is not symmetric, wave conversion occurs; for example, at the discontinuities bending waves are converted to longitudinal waves and vice versa. A consequence of that is that within the stopping bands the vibration decay may be rather pronounced near the point of excitation and much smaller in some distance. The reason for this phenomenon is that after a certain distance the wave type with the smallest decay rate is responsible for the transport of energy.[2, 14]

(d) The resonance frequencies of finite periodic structures always lie within the passing bands.

(c) For infinite periodic structures the input impedance for point or moment excitation is purely imaginary within the stopping bands.

APPENDIX: SOME EQUATIONS RELATED TO PLANE ONE- AND TWO-DIMENSIONAL STRUCTURES WITHOUT BOUNDARIES

A.1 Torsional Waves in Rods

1. *Equation of Motion*

$$GI \frac{\partial^2 w}{\partial x_1^2} - \rho\Theta \frac{\partial^2 w}{\partial t^2} = 0[= -M_d(x_1)],$$

where GI is torsional stiffness, Θ the moment of inertia (both per unit length), G the shear modulus, w the angle of torsion, and M_d the driving moment. For circular cross section $\Theta = \frac{1}{2}\pi(a_1^4 - a_2^4)$; $I = \frac{1}{2}\pi(a_1^4 - a_2^4)$. For rectangular cross section $\Theta = \frac{1}{12}(bh^3 + hb^3)$; $I = \alpha b^3 h$ (for $h = b$, $\alpha = 0.141$; for $h = 2b$, $\alpha = 0.229$; for $h = 10b$, $\alpha = 0.312$).

The following equations are equivalent:

$$M = -GI \frac{\partial w}{\partial x_1}, \qquad -\rho\Theta \frac{\partial^2 w}{\partial t^2} \, [+M_d(x_1)] = \frac{\partial m}{\partial x_1}.$$

2. *Dispersion Relation*

$$GIk_{1f}^2 = \rho\Theta\omega^2, \qquad c_T = (GI/\rho\Theta)^{1/2}.$$

3. *Energy Density*

$$W_{e,\text{kin}} = \frac{1}{2}\rho\Theta w^2, \qquad W_{e,\text{pot}} = \frac{1}{2} GI \left(\frac{\partial w}{\partial x_1} \right)^2.$$

4. *Moment Impedance Semi-infinite, Excited at One End*

$$Z_M = \frac{M_d}{w(0)} = \sqrt{\rho\Theta GI} = \rho\Theta c_T.$$

A.2 Bending Waves in Beams (Bernoulli–Euler)

1. *Equation of Motion (See Fig. 1c)*

$$B \frac{\partial^4 \xi_2}{\partial x_1^4} + \rho S \frac{\partial^2 \xi_2}{\partial t^2} = p_2,$$

where $B = ES\kappa^2$ is the bending stiffness, S the cross section, κ the radius of gyration (rectangular: $\kappa = h/\sqrt{12}$; circular with outer radius a_1 and inner radius a_2: $\kappa = 0.5(a_1^2 + a_2^2)^{1/2}$.

The following first-order equations are equivalent:

$$w = \frac{\partial \xi_2}{\partial x_1}, \qquad M = B\,\frac{\partial^2 \xi_2}{\partial x_1^2}, \qquad F_2 = \frac{\partial M}{\partial x_1},$$

$$\frac{\partial F_2}{\partial x} = -\rho S\,\frac{\partial^2 \xi_2}{\partial t^2},$$

where w is the bending angle, M the bending moment, F_2 the force in the x_2-direction.

2. *Dispersion Relation*

$$Bk_B^4 = \omega^2 \rho S, \qquad k_B = \omega / c_B = \left(\frac{\omega^2 \rho S}{B}\right)^{1/4}.$$

Group velocity $c_g = 2c_B$,

3. *Energy Density*

$$W_{e,\text{kin}} = \frac{1}{2}\,\rho S \dot{\xi}_2^2, \qquad W_{e,\text{pot}} = \frac{1}{2}\,B\left(\frac{\partial^2 \xi_2}{\partial x_1^2}\right)^2.$$

4. *Point Impedance*

$$Z_F = \begin{cases} \dfrac{2\rho S\omega(1+j)}{k_B} & \text{(infinite, excited in the middle),} \\[2ex] \dfrac{\rho S\omega(1+j)}{2k_B} & \text{(semifinite, excited at end).} \end{cases}$$

A.3 In-plane Waves in Isotropic Plates

1. *Equations of Motion*

$$\left(\frac{2}{1-\mu}\,\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_3^2} - \frac{\rho}{G}\,\frac{\partial^2}{\partial t^2}\right)\xi_1 + \frac{1+\mu}{1-\mu}\,\frac{\partial^2 \xi_3}{\partial x_1\,\partial x_3}$$
$$= \frac{-1}{G}\left[\frac{\mu}{2(1-\mu)}\,\frac{\partial p_2}{\partial x_1} + \frac{p_1}{h}\right],$$

$$\frac{1+\mu}{1-\mu}\,\frac{\partial^2 \xi_1}{\partial x_1\,\partial x_3} + \left(\frac{\partial^2}{\partial x_1^2} + \frac{2}{1-\mu}\,\frac{\partial^2}{\partial x_3^2} - \frac{\rho}{G}\,\frac{\partial^2}{\partial t^2}\right)\xi_3$$
$$= \frac{-1}{G}\left[\frac{\mu}{2(1-\mu)}\,\frac{\partial p_2}{\partial x_3} + \frac{p_3}{h}\right],$$

where p_2 is the force per unit area acting in x_2-direction (normal to plate), p_1 and p_3 the forces per unit area acting in the x_1-, and x_3-direction (parallel to the plate), G the shear modulus, and μ Poisson's ratio.

2. *Dispersion Relations*

$$k_{1S}^2 = \frac{\omega^2 \rho}{G} - k_3^2 \quad \text{(shear waves)},$$

$$k_{1L}^2 = \frac{\omega^2 \rho}{2G}\,(1-\mu) - k_3^2 \quad \text{(longitudinal waves)}.$$

The plane-wave solution (for k_3 arbitrary) is given as

$$\xi_1(x_1, x_3) = [A_{S+}e^{-jk_{1S}x_1} + A_{S-}e^{jk_{1S}x_1} + A_{L+}e^{-jk_{1L}x_1}$$
$$+ A_{L-}e^{jk_{1L}x_1}]e^{-jk_3x_3}$$

$$\xi_3(x_1, x_3) = \left[-A_{S+}\,\frac{k_{1S}}{k_3}\,e^{-jk_{1S}x_1} + A_{S-}\,\frac{k_{1S}}{k_3}\,e^{jk_{1S}x_1}\right.$$
$$\left. + A_{L+}\,\frac{k_3}{k_{1L}}\,e^{-jk_{1L}x_1} - A_{L-}\frac{k_3}{k_{1L}}\,e^{jk_{1L}x_1}\right]$$
$$\cdot\, e^{-jk_3x_3}.$$

3. *Energy Density*

$$W_{e,\text{kin}} = \frac{1}{2}\,\rho h(\dot{\xi}_1^2 + \dot{\xi}_3^2),$$

$$W_{e,\text{pot}} = \frac{Eh}{2(1-\mu^2)}\left[\left(\frac{\partial \xi_1}{\partial x_1}\right)^2 + \left(\frac{\partial \xi_3}{\partial x_3}\right)^2\right.$$
$$+ 2\mu\,\frac{\partial \xi_1}{\partial x_3}\,\frac{\partial \xi_3}{\partial x_1}$$
$$\left. + \frac{1-\mu}{2}\left(\frac{\partial \xi_1}{\partial x_3} + \frac{\partial \xi_3}{\partial x_1}\right)^2\right].$$

4. *Point Impedance*[15]

$$\frac{1}{Z_F} = \frac{\omega(3-\mu)}{16hG}$$
$$+ j\,\frac{\omega}{8\pi hG}\left[(1-\mu)\ln\frac{2c_L}{\omega a} + 2\ln\frac{2c_S}{\omega a}\right],$$

$$c_S = \sqrt{\frac{G}{\rho}}, \qquad c_L = \sqrt{\frac{E}{[(1-\mu^2)\rho]}}.$$

A.4 Bending Waves in Isotropic Plates Including Shear Stiffness and Rotatory Inertia (Timoshenko–Mindlin Theory[2,3]

1. *Equation of Motion*

$$D\Delta\Delta\xi_2-\left(\frac{D\rho}{G}+\rho I\right)\Delta\frac{\partial^2\xi_2}{\partial t^2}+\rho h\frac{\partial^2\xi_2}{\partial t^2}$$
$$+\rho I\frac{\rho}{G}\frac{\partial^4\xi_2}{\partial t^4}$$
$$=p_2-\frac{D}{Gh}\Delta p_2+\frac{\rho I}{Gh}\frac{\partial^2 p_2}{\partial t^2},$$

where D is the plate bending stiffness, G the shear modulus, $I = h^3/12$, and p_2 the force per unit area in the x_2-direction (normal to plate).

An equivalent form is

$$Gh(\Delta\xi_2+\Psi)=\rho h\frac{\partial^2\xi_2}{\partial t^2}-p_2,$$

$$D\Delta\Psi-Gh(\Delta\xi_2+\Psi)=\rho I\frac{\partial^2\Psi}{\partial t^2},$$

$$\Psi=\frac{\partial w_1}{\partial x_1}+\frac{\partial w_3}{\partial x_3},$$

where w_1 and w_3 are bending angles.

2. *Dispersion Relation*

$$k_1^2=(k_{1f}^2+k_{3f}^2)_I=\beta_{SL}^2+\left(\beta_{SL}^4+\frac{\alpha\omega^2\rho h}{D}\right)^{1/2}$$

(propagating),

$$k_{1I}^2=(k_{1f}^2+k_{3f}^2)_{1I}=\beta_{SL}^2-\left(\beta_{SL}^4+\frac{\alpha\omega^2\rho h}{D}\right)^{1/2}$$

(near field),

with

$$\beta_{SL}^2=\frac{\omega^2}{2}\left(\frac{1}{c_T^2}+\frac{1}{c_L^2}\right),\qquad \alpha=1-\frac{\omega^2h^2}{12G}$$

and $c_T = (G/\rho)^{1/2}$ is the shear wave speed and $c_L = (D/\rho I)^{1/2} = (E/\rho(1-\mu^2)^{1/2}$ the longitudinal wave speed. For $\alpha > 0$ there are propagating waves and near fields; for $\alpha < 0$ there are only propagating waves.

3. *Energy Density*

$$W_{e,\mathrm{kin}}=\frac{\rho h}{2}\left[\dot\xi_2^2+\frac{h^2}{12}(\dot w_1^2+\dot w_3^2)\right],$$

$$W_{e,\mathrm{pot}}=\frac{E}{2(1-\mu^2)}\frac{h^3}{12}\left[\left(\frac{\partial w_1}{\partial x_1}\right)^2+\left(\frac{\partial w_3}{\partial x_3}\right)^2\right.$$
$$\left.+2\mu\left(\frac{\partial w_1}{\partial x_3}\frac{\partial w_3}{\partial x_1}\right)\right]$$
$$+\frac{Gh}{2}\left[\left(w_1+\frac{\partial\xi_2}{\partial x_1}\right)^2+\left(w_3+\frac{\partial\xi_2}{\partial x_3}\right)^2\right.$$
$$\left.+\frac{h^2}{12}\left(\frac{\partial w_1}{\partial x_3}+\frac{\partial w_3}{\partial x_1}\right)^2\right].$$

4. *Point Impedance*

$$Z_F=\frac{8D}{\omega}\frac{k_I^2-\beta_{SL}^2}{A_R+jA_I},$$

$$A_R=\begin{cases}\alpha+2\dfrac{D}{Gh}(k_I^2-\beta_{SL}^2) & \text{for}\,\alpha>0,\\[2ex] 2\dfrac{D}{Gh}(k_I^2-\beta_{SL}^2) & \text{for}\,\alpha<0,\end{cases}$$

$$A_1=\frac{-2}{\pi}\left[\alpha\ln\left(\frac{\omega}{k_I^2}\sqrt{\frac{\rho h|\alpha|}{D}}\right)\right.$$
$$\left.+\frac{D}{Gh}(k_I^2\ln k_Ia-k_{II}^2\ln k_{II}a)\right],$$

where a is the radius of the excited area.

A.5 Bending Waves in Orthotropic Plates (Bernoulli–Euler)

1. *Equation of Motion*[16]

$$D_1\frac{\partial^4\xi_2}{\partial x_1^4}+2(D_\mu+2D_G)\frac{\partial^4\xi}{\partial x_1^2\partial x_3^2}$$
$$+D_3\frac{\partial^4\xi_2}{\partial x_3^4}+\rho h\frac{\partial^2\xi_2}{\partial t^2}=p_2,$$

where D_1, D_3 are the bending stiffness in the x_1- and x_3-direction. For anisotropic materials $D_1 = E_1h^3/12(1-\mu^2)$, $D_3 = E_3h^3/12(1-\mu^2)$, $D_\mu = \frac{1}{12}E_\mu h^3$, and $D_G = \frac{1}{12}Gh^3$.

2. *Dispersion Relation*

$$D_1 k_{1f}^4 + 2(D_\mu + 2D_G) k_{1f}^2 k_{3f}^2 + D_3 k_{3f}^4 - \omega^2 \rho h = 0.$$

Here, the angle dependence is obtained by setting $k_{1f} = k_f \sin \vartheta$; $k_{3f} = k_f \cos \vartheta$.

3. *Point Impedance*

$$Z_F = \frac{4\pi}{K(\alpha)} (\rho^2 h^2 D_1 D_3)^{1/4},$$

$$\alpha = \left\{ \frac{1}{2} \left[1 - \frac{D_\mu + 2D_G}{(D_1 D_3)^{1/2}} \right] \right\}^{1/2},$$

where $K(\alpha)$ is the complete elliptic integral of the first kind. Useful approximations are $D_\mu + 2D_G \approx (D_1 D_3)1/2$ and $K(\alpha) = \frac{1}{2}\pi$.

REFERENCES

1. Lord Rayleigh, *The Theory of Sound*, Macmillan, London, 1877.
2. L. Cremer, M. Heckl, and E. E. Ungar, *Structure-Borne Sound*, Springer, Berlin, 1988.
3. M. C. Junger and D. Feit, *Sound Structures and Their Interaction*, 2nd ed., MIT Press Cambridge, MA.
4. F. J. Fahy, *Sound and Structural Vibration*, Academic, London, 1989.
5. L. Cremer, "Theorie der Schalldämmung dünner Wände bei schrägem Einfall," *Akustische Zeitschrift*, Vol. VII, 1942, p. 81.
6. L. L. Vér, "Interaction of Sound Waves with Solid Structures," L. L. Beranek, I. L. Vér (Eds.), *Noise and Vibration Control Engineering*, Wiley, New York, 1992, Chapter 9.
7. J. Scheuren, "Application of the Method of Integral Equations to the Vibrations of Plates," *Recent Advances in Structural Dynamics*, Vol. I, ISVR, University of Southampton, Southampton, 1984, pp. 171–177.
8. O. L. Zienkiewicz, *The Finite Element Method*, McGraw-Hill, London, 1977.
9. R. D. Blevins, *Formulas for Natural Frequency and Mode Shape*, Van Nostrand, New York, 1979.
10. A. W. Leissa, "Vibration of Plates," NASA SP-160, Washington, DC, 1969.
11. A. W. Leissa, "Vibration of Shells," NASA SP-288, Washington, DC, 1973.
12. L. Brillouin, *Wave Propagation in Periodic Structures*, Dover, New York, 1953.
13. D. J. Mead, "Response of Periodic Structures to Noise Fields," in R. G. White (Ed.), *Noise and Vibration*, Ellis Horwood, Chichester, 1982, pp. 285–306.
14. V. D. Belov, S. A. Rybak, B. A. Tartakovskii, "Propagation of Vibrational Energy in Absorbing Structures," *Sov. Phys. Acoust.*, Vol. 23, 1977, pp. 115–118.
15. S. Ljunggren, "Generation of Waves in an Elastic Plate by a Torsional Moment and a Horizontal Force," *J. Sound Vib.*, Vol. 93, 1984, pp. 161–187.
16. M. Heckl, "Untersuchungen an orthotropen Platten," *Acustica*, Vol. 10, 1960, pp. 109–115.

65

NONLINEAR VIBRATION

LAWRENCE N. VIRGIN AND EARL H. DOWELL

1 INTRODUCTION

Problems in nonlinear mechanical vibration are considerably more difficult to analyze than their linear counterparts. The presence of nonlinear forces in many physical systems results in a spectacular array of possible behavior even for relatively low-order models. Mechanisms of instability and sudden changes in response are of fundamental importance in the behavior of nonlinear systems. The closed-form solutions familiar from linear vibration theory are of limited value and give little hint as to the complexity and sometimes unpredictability of typical nonlinear behavior.

Solutions to the governing nonlinear equations of motion are obtained either using approximate analytical methods or numerical simulation guided by dynamical systems theory. Due to the inherent complexity of nonlinear vibration, the use of the geometric, qualitative theory of ordinary differential equations play a central role in the classification and analysis of such systems.

This chapter is divided into two main sections. The first describes the effects of nonlinear stiffness and damping on free oscillations resulting in multiple equilibria, amplitude-dependent frequencies, basins of attraction, and limit cycle behavior. The second describes the effects of external periodic forcing resulting in nonlinear resonance, hysteresis, subharmonics, and chaos. Low-order, archetypal examples with relevance to mechanical and electrical oscillators are used to illustrate the phenomena.

2 NONLINEAR FREE VIBRATION

2.1 Autonomous Dynamical Systems

Nonlinear free vibration problems are generally governed by sets of n first-order ordinary differential equations of the form[1]

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}), \tag{1}$$

where the function $\mathbf{f}$ defines a vector field in the n-dimensional phase space, and an overdot represents a time derivative. In single-degree-of-freedom oscillators position and velocity are the two state variables. Often the stiffness terms are a function of displacement only and can be derived from a potential energy function. Many mechanical systems have the property that for small displacements this function will be linear in x, and furthermore the damping is often assumed to be governed by a simple, viscous energy dissipation giving linear terms in $\dot{x}$. In this case standard vibration theory (see Chapter 63) can be used to obtain exact closed-form solutions giving an exponentially decaying (sinusoidal or monotonic) response.[2,3] Linear algebra techniques and modal analysis can be used for multi-degree-of-freedom systems. For systems where the induced nonlinearity is small, linearization and certain approximate analytical techniques can be used, but if no restriction is placed on nonlinearity, then numerical methods must generally be used to solve Eq. (1). Specific examples will illustrate typical transient behavior.

2.2 Simple Pendulum

Small-Amplitude Linear Behavior As an introduction to typical nonlinear behavior, consider the simple

---

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker
Copyright © 1997 John Wiley & Sons, Inc.

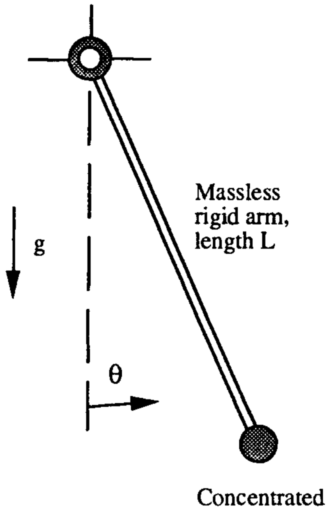


Fig. 1 Schematic diagram of the simple rigid-arm pendulum.

rigid-arm pendulum shown schematically in Fig. 1. It is a simple matter, using Newton's laws or Lagrange's equations,[4] to derive the governing equation of motion

$$\ddot{\theta} + b\dot{\theta} + \frac{g}{L}\sin\theta = 0, \qquad (2)$$

where b is the (viscous) damping coefficient. Note that the linear, undamped natural frequency is a constant $\omega_0 = \sqrt{g/L}$. For small angles, $\sin\theta \cong \theta$, and given the two initial conditions θ_0 and $\dot{\theta}_0$, the pendulum will undergo periodic motion that decays with a constant (damped) period as $t \to \infty$, coming to rest at $\theta = 0$, the position of static stable equilibrium (Fig. 2*a*). This equilibrium position acts as a point attractor for all local transients.

It is very useful in nonlinear vibrations to look at phase trajectories in a plot of displacement against velocity. This is shown in Fig. 2*b* for the pendulum started from rest at $\theta_0 = \pi/8$. The near elliptical nature of these curves (indicating sinusoidal motion) is apparent as the motion evolves in a clockwise direction.

The motion of the pendulum can be thought to be occurring within the potential energy well shown in Fig. 3, that is, the integral of the restoring force. The assumption of small angles effectively means that the restoring force is assumed to be linear about the origin (Hooke's law), with a locally parabolic potential energy well. Using the analogy of a ball rolling on this surface, it is easy to visualize the motion shown in Fig. 2.

Large-Amplitude Nonlinear Behavior For large-angle swings of the pendulum the motion is no longer linear. The restoring force induces a softening spring effect. The natural frequency of the system will now be a function of amplitude as the unstable equilibrium at $\theta = \pi$ comes into effect. Linear theory can also be used about an unstable equilibrium, and local trajectories will tend to diverge away from a saddle point. Physically this is the case of the pendulum balanced upside down. The process of linearization involves the truncation of a power series about the equilibrium positions and characterizes the local motion. Often this information can then be pieced together to obtain a qualitative impression of the complete phase portrait.[5]

Consider the undamped ($b = 0$) pendulum started from rest at $\theta(0) = 0.99\pi$ and shown as a velocity time series in Fig. 4*a* based on direct numerical simulation. These oscillations are far from sinusoidal. The pendulum slows down as it approaches its inverted position, with a natural period of approximately 3.5 times the linear natural

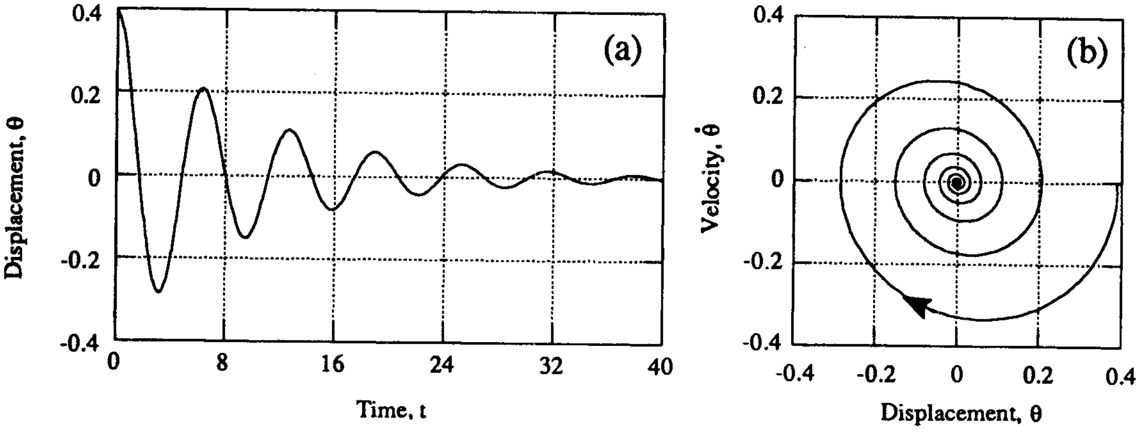


Fig. 2 Small-amplitude swings of the damped pendulum illustrating linear motion: (*a*) time series; (*b*) phase portrait ($\theta_0 = \pi/8$, $\omega_0 = 1$, $b = 0.1$).

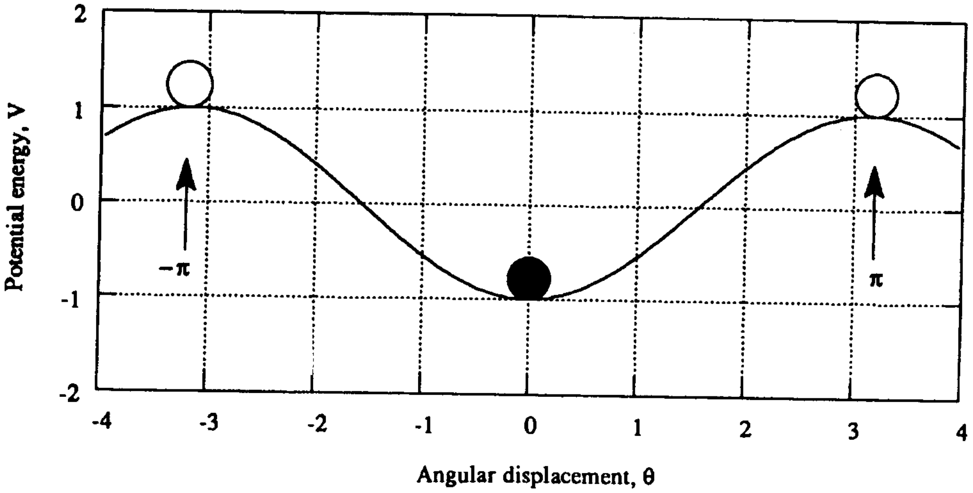


Fig. 3 Underlying potential energy for the pendulum showing the analogy of a rolling ball. The black ball represents stable equilibrium.

period of 2π. Exact analytical solutions are available here in the form of elliptic integrals,[6] and again this motion can be illustrated as a phase portrait (Fig. 4*b*). Since the phase space for this system is the surface of a cylinder, the saddle points corresponding to the hilltops in Fig. 3 represent the same inverted position. Here the phase trajectories (for this undamped system) can be considered as contours of constant total energy, that is,

$$E = \tfrac{1}{2}\dot{\theta}^2 - \omega_0^2 \cos\theta, \tag{3}$$

where the initial conditions $(\theta_0, \dot{\theta}_0)$ determine E. The phase trajectories described by Eq. (3) will be closed (i.e., periodic) for relatively small initial conditions, but open orbits corresponding to a spinning pendulum result if the starting conditions provide sufficient initial energy.

2.3 Duffing's Equation

Another archetypal example of an oscillator with a nonlinear restoring force is an autonomous form of Duffing's equation:

$$\ddot{x} + b\dot{x} + \alpha x + \beta x^3 = 0, \tag{4}$$

which can, for example, be used to study the large-amplitude motion of a pre- or postbuckled beam[4] or plate or the moderately large-amplitude motion of the pendulum. The free vibration of a beam loaded axially beyond

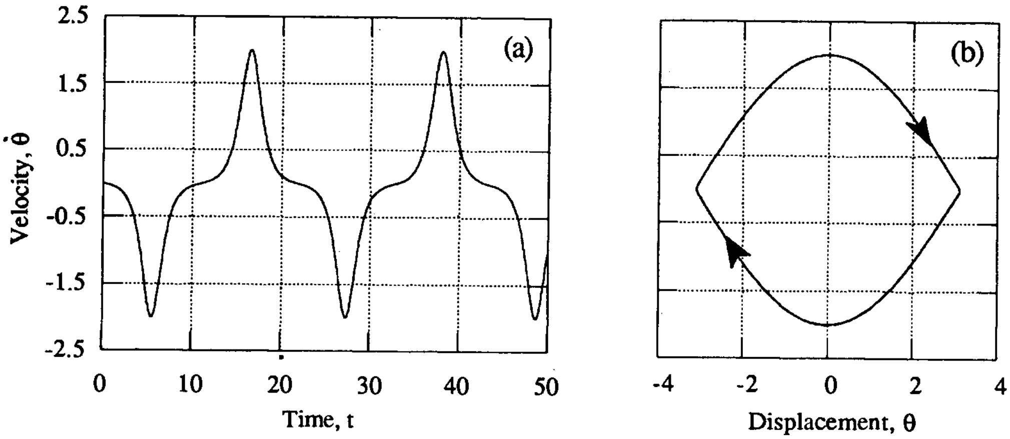


Fig. 4 Large undamped swings of the undamped pendulum: (*a*) velocity time series; (*b*) phase portrait ($\theta_0 = 0.99\pi$, $\omega_0 = 1$).

its elastic critical limit can be modeled by Eq. (4) with negative linear and positive cubic stiffness. In this case the origin (corresponding to the straight configuration) is unstable with two stable equilibrium positions at $x = \pm 2$, for example, when $\alpha = -1$ and $\beta = 4$. Now, not only does the natural period depend on initial conditions but so does the final resting position as the two stable equilibria compete to attract transients. Each equilibrium is surrounded by a domain of attraction. These interlocking domains, or catchment regions, are defined by the separatrix, which originates at the saddle point. It is apparent that some transients will traverse the (double) potential energy well a number of times before sufficient energy has been dissipated and the motion is contained within one well. Clearly it is difficult to predict which long-term resting position will result given a level of uncertainty in the initial conditions.[4,7]

2.4 Van der Pol's Oscillator

In the previous section the nonlinearity in the stiffness resulted in behavior dominated by point attractors (equilibria). Periodic attractors (limit cycles) are also possible in autonomous dynamical systems with nonlinear damping. A classical example relevant to electric circuit theory is Van der Pol's equation,[1,8]

$$\ddot{x} - h(1 - x^2)\dot{x} + x = 0, \tag{5}$$

where h is a constant. Here, the damping term is positive for $x^2 > 1$ and negative for $x^2 < 1$, for a positive h. When h is negative, local transient behavior simply decays to the origin rather like in Fig. 2. However, for positive h, a stable limit cycle behavior occurs where the origin is now unstable and solutions are attracted to a finite-amplitude steady-state oscillation. This is shown in Figure 5, for $h = 1$ and $x_0 = \dot{x}_0 = 0.1$, as a time series in 5*a* and a phase portrait in 5*b*. Initial conditions on the outside of this limit cycle would be similarly attracted. This phenomenon is also known as a self-excited or relaxation oscillation.[8,9] Related physical examples include mechanical chatter due to dry friction effects between a mass and a moving belt (stick-slip)[10] and certain flow-induced aeroelastic problems including galloping and flutter.[11]

2.5 Instability

Nonlinear free vibration problems are generally dominated by the influence of equilibria on transient behavior. There are several definitions of stability, but generally if a small perturbation causes the system to move away from equilibrium, then this is a locally unstable state. This is familiar from linear vibration theory where transient behavior is described by the characteristic eigenvalues of the system. For example, the linear oscillator in Fig. 2 has two complex conjugate characteristic eigenvalues with negative real parts. A linearization about the inverted position for the pendulum would lead to one positive real eigenvalue resulting in divergence. Similarly, behavior in the vicinity of equilibrium for the Van der Pol oscillator ($h > 0$) is characterized by complex conjugate eigenvalues but with positive real parts and hence a growth of unstable oscillations. The amplitude of the oscillation in this case is limited by nonlinear effects.

Such systems tend to be sensitive to certain control parameters, for example, h in Eq. (5). More generally a system of the form

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mu) \tag{6}$$

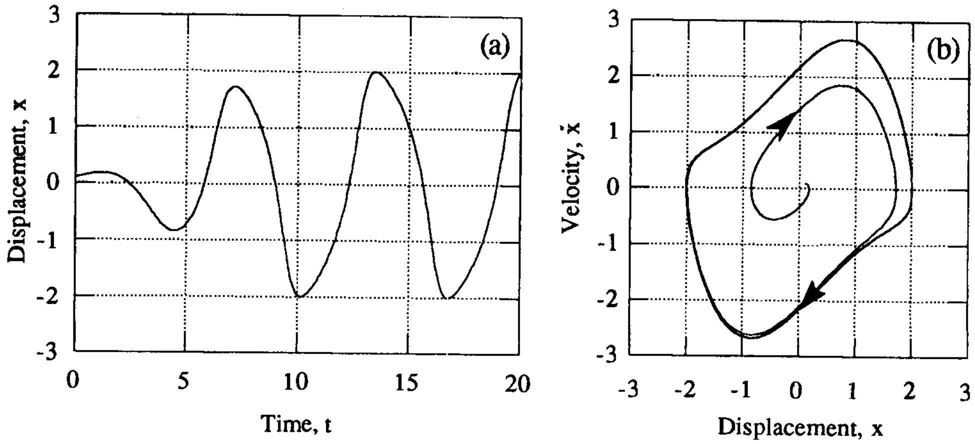


Fig. 5 Stable limit cycle behavior exhibited by Van der Pol's equation: (*a*) time series; (*b*) phase portrait ($x_0 = 0.1$, $\dot{x}_0 = 0.1$, $h = 1.0$).

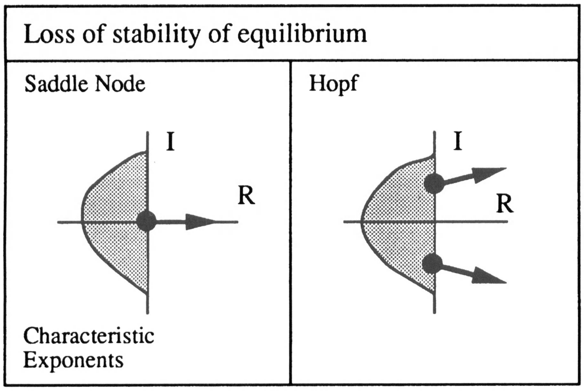


Fig. 6 Two generic instabilities of equilibria under the operation of a single control.

must be considered where the external parameters $\boldsymbol{\mu}$ allow a very slow (quasi-static) evolution of the dynamics. In nonlinear dynamic systems such changes in these parameters may typically result in a qualitative change (bifurcation) in behavior, and for systems under the operation of one control parameter, these instabilities take place in one of two well-defined, generic ways. Both of these instability mechanisms have been encountered in the previous sections. The loss of stiffness is associated with the passage of an eigenvalue in to the positive half of the complex plane as shown in Fig. 6. This stability transition is known as the saddle-node but is also referred to as a fold bifurcation and is encountered, for example, in limit-point snap-through buckling in structural mechanics.[4] The loss of damping is associated with the passage of a pair of complex conjugate eigenvalues in to the positive half of the complex plane. This instability mechanism is known as a Hopf bifurcation.[12,13]

3 NONLINEAR FORCED VIBRATION

3.1 Nonautonomous Dynamical Systems

When a dynamic system is subjected to external excitation, a mathematical model of the form

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t) \tag{7}$$

results. This system can be made equivalent to Eq. (1) by including the dummy equation $\dot{t} = 1$ to give an extra phase coordinate. In forced nonlinear vibration primary interest is focused on the case where the input (generally force) is assumed to be a periodic function. For a single-degree-of-freedom oscillator, the forcing phase effectively becomes the third state variable. If the system is linear, then the solution to Eq. (7) will consist of a superposition of a complementary function, which governs the transient free-decay response, and a particular integral, which governs the steady-state forced response (see Chapter 63). The steady-state oscillation generally responds with the same frequency as the forcing and acts as a periodic attractor to local transient behavior.

If Eq. (7) is nonlinear, then it is generally not possible to obtain an exact analytical solution. A number of choices are available: (i) linearize about the static equilibria and obtain solutions that are valid only in a local sense, (ii) use an approximate analytical method and assume relatively small deviation from linearity, for example, use a perturbation scheme,[14] (iii) simulate the equation of motion using numerical integration.[15] The first of these approaches leads to the familiar resonance results from standard vibration theory (see Chapter 63). A large body of research has been devoted to the second approach[1,14,16,17]. However, due to recent advances in, and availability of, digitial computers and sophisticated graphics, the third approach has achieved widespread popularity in the study of nonlinear vibrations. Furthermore the numerical approach has been enhanced by the development of the qualitative insight of dynamical systems theory.[18]

3.2 Nonlinear Resonance

The steady-state response of nonlinear vibration problems modeled by Eq. (7) with external forcing of the form

$$G \sin(\omega t + \phi) \tag{8}$$

exhibit some interesting differences from their linear counterparts. Consider the harmonically forced Duffing equation, that is, Eq. (8) added to the right-hand side of Eq. (4). For sufficiently large G the cubic nonlinearity is induced and a typical resonance amplitude response is plotted as a function of forcing frequency in Fig. 7 for $G = 0.1$ and three different damping levels. These curves were obtained using the method of harmonic balance[1,7] and illustrate several nonlinear features. For relatively small amplitudes, that is, in a system with relatively small external forcing, or heavy damping, and generally away from resonance, the response is not significantly different from the linear case. However, for larger amplitudes the softening spring effect causes a bending over of the curves. This is related to the lengthening effect of amplitude on the natural period of the underlying autonomous system and causes the maximum amplitude to occur somewhat below the natural frequency, and for some frequencies multiple solutions are possible. For example, for the case $b = 0.18$, there are three steady-state solutions near $\omega = 0.8$. Two are stable and are separated by an unstable solution (not shown).

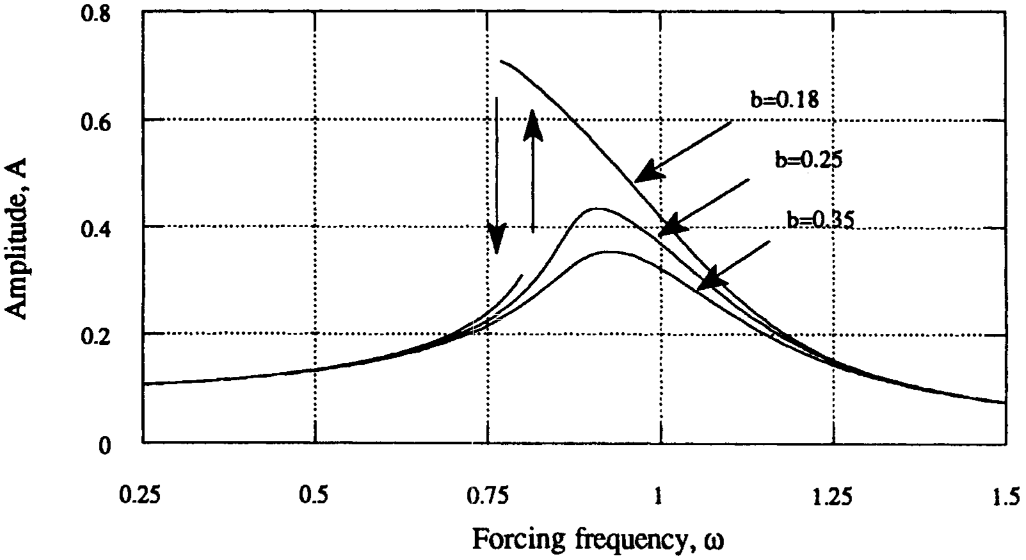


Fig. 7 Nonlinear resonance illustrating the jump in amplitude. This steady-state motion occurs about the offset static equilibrium position (b = 0.18, 0.25, 0.35, $\alpha = -1$, $\beta = 1$, $G = 0.1$).

The two stable steady-state cycles are periodic attractors and compete for the capture of transient trajectories, that is, in this region different initial conditions lead to different persistent behavior. The jump phenomenon is observed by gradually (quasi-statically) changing the forcing frequency ω while keeping all the other parameters fixed. Starting from small ω and slowly increasing, the response will suddenly exhibit a finite jump in amplitude and follow the upper path. Now starting with a large ω and gradually reducing, the response will again reach a critical state resulting in a sudden jump down to the small-amplitude solution. These jumps occur at the points of vertical tangency and bound a region of hysteresis.

A hardening spring exhibits similar features except the resonance curves bend to the right. A physical example of this type of behavior can be found in the large-amplitude lateral oscillations of a thin elastic beam or plate. In this case the induced in-plane (stretching) force produces a hardening (cubic) nonlinearity in the equations of motion, that is, α and β both positive, and for certain forcing frequencies the beam can be perturbed from one type of motion to another. The domains of attraction often consist of interlocking spirals, and hence it may be difficult to determine which solution will persist given initial conditions to a finite degree of accuracy in a similar manner to the unforced case.

3.3 Subharmonic Behavior

Periodic solutions to ordinary differential equations need not necessarily have the same period as the forcing term. It is possible for a system to exhibit subharmonic behavior, that is, the response has a period of n times the forcing period. Subharmonic motion can also be analyzed using approximate analytical techniques.[1,14,16]

Figure 8 illustrates a typical subharmonic of order 2 obtained by numerically integrating Duffing's equation:

$$\ddot{x} + b\dot{x} + \alpha x + \beta x^3 = G \sin \omega t \tag{9}$$

for $b = 0.3$, $\alpha = -1$, $\beta = 1$, $\omega = 1.2$, and $G = 0.65$. Subharmonic behavior is generally a nonlinear feature and is often observed when the forcing frequency is close to an integer multiple of the natural frequency. Also, subharmonics may result due to a bifurcation from a harmonic response, and thus may complicate the approximate scenario shown in Fig. 7. Superharmonics corresponding to a response that repeats itself a number of times within each forcing cycle are also possible especially at low excitation frequencies.

3.4 Poincaré Sampling

A useful qualitative tool in the analysis of periodically forced nonlinear systems is the Poincaré section.[5] This technique consists of stroboscopically sampling the trajectory every forcing cycle or at a defined surface of section in the phase space. The accumulation of points will then reflect the periodic nature of the response. A fundamental harmonic response appears as a single point simply repeating itself. The location of this point on the phase trajectory and hence in the $(x, \dot{x})$ projection depends on the initial conditions. For example, consider

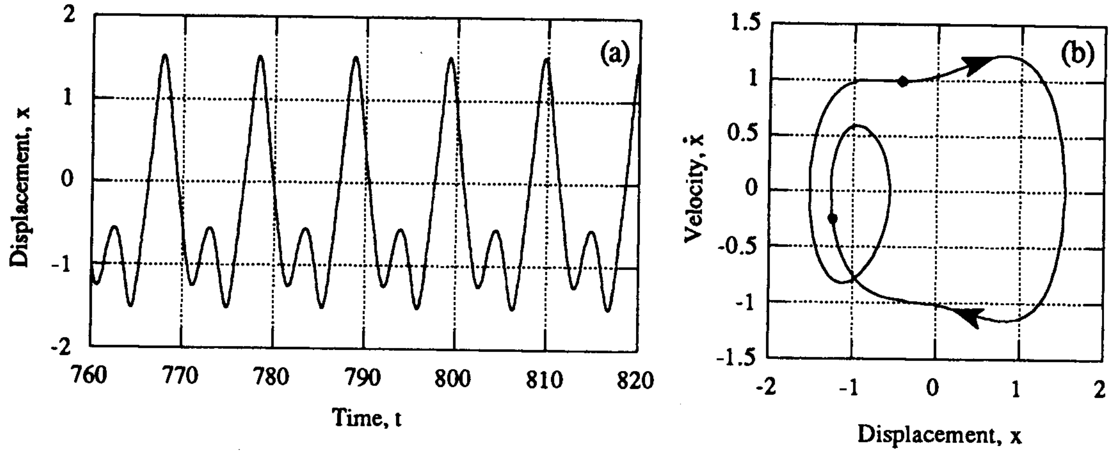


Fig. 8 Subharmonic oscillations in Duffing's equation: (*a*) time series; (*b*) phase projection ($b = 0.3$, $\alpha = -1$, $\beta = 1$, $\omega = 1.2$, $G = 0.65$).

the subharmonic response of Fig. 8. The forcing period is $T = 2\pi/\omega = 5.236$, and if the response is inspected whenever t is a multiple of T then two points are visited alternately by the trajectory, as shown by the dots in Fig. 8*b*. The Poincaré sampling describes a mapping and can also be used to study the behavior of transients, and hence stability. Further refinements to this technique have been developed including the reconstruction of attractors using time-delay sampling.[5] This is especially useful in experimental situations where there may be a lack of information about certain state variables or in autonomous systems where there is no obvious sampling period.

3.5 Quasi-periodicity

Another possible response in forced nonlinear vibration problems is the appearance of quasi-periodic behavior. Although the superposition of harmonics, including the beating effect, is commonly encountered in coupled linear oscillations, it is possible for nonlinear single-degree-of-freedom systems to exhibit a relatively complicated response where two or more incommensurate frequencies are present. For a two-frequency response local transients are attracted to the surface of a toroidal phase space, and hence Poincaré sampling leads to a continuous closed orbit in the projection because the motion never quite repeats itself. Although a time quasi-periodic time series may look complicated it is predictable. The fast Fourier transform (FFT) is a useful technique for identifying the frequency content of a signal and hence distinguishing quasi-periodicity from chaos.

3.6 Piecewise Linear Systems

There are many examples in mechanical engineering where the stiffness of the system changes abruptly, for example, where a material or component acts differently in tension and compression, or a ball bouncing on a surface. Although they are linear within certain regimes, the stiffness is dependent on position, and their behavior is often strongly nonlinear. If such systems are subjected to external excitation, then a complex variety of responses are possible.[19,20] Related problems include the backlash phenomenon in gear mechanisms where a region of free-play exists.[4] Nonlinearity also plays a role in some coulomb friction problems including stick-slip where the discontinuity is in the relative velocity between two dry surfaces.[1] Feedback control systems often include this type of nonlinearity.[21]

3.7 Chaotic Oscillations

The possibility of relatively simple nonlinear systems exhibiting extremely complicated (unpredictable) dynamics was known to Poincaré about 100 years ago and must have been observed in a number of early experiments on nonlinear systems. However, the relatively recent ability to simulate highly nonlinear dynamical systems numerically has led to a number of interesting new discoveries with implications for a wide variety of applications, especially chaos, that is, a fully deterministic system that exhibits randomlike behavior and an extreme sensitivity to initial conditions. This has profound implications for the concept of predictability and has stimulated intensive research.[22]

Figure 9 shows a chaotic time series obtained by the numerical integration of Eq. (9) for appropriately selected b, α, β, G, and ω using a fourth-order Runge–Kutta scheme.[14] Here, transient motion is allowed to decay leaving the randomlike wave form shown (Fig. 9*a*). This response, which traverses both static equilibria ($x_e = \pm 1$), may be considered to have an infinite period. An important feature of chaos, in marked

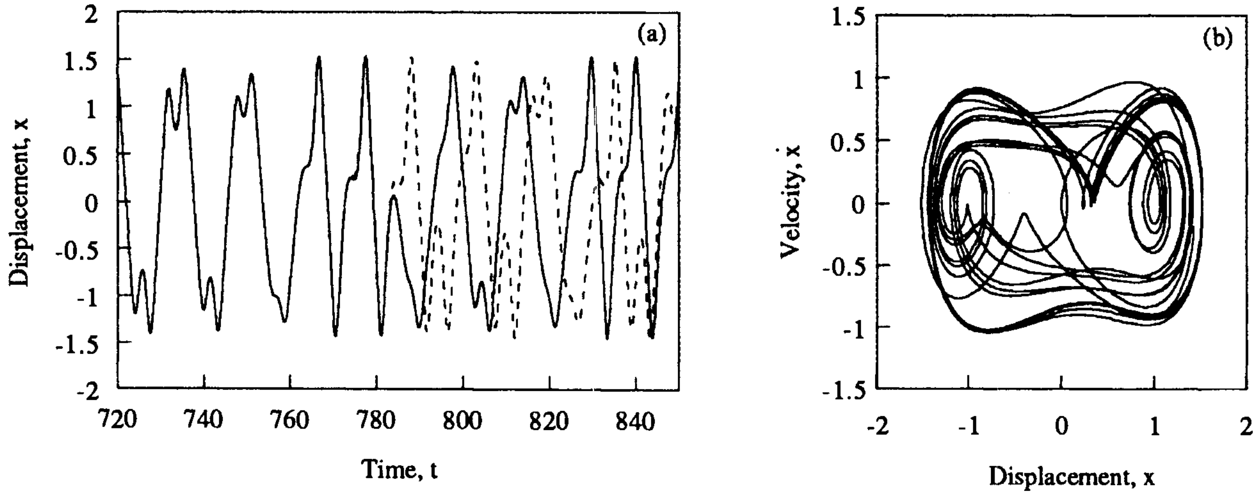


Fig. 9 Chaotic response of Duffing's equation including exponential divergence of initially close trajectories, parameters are the same as Fig. 8, $G = 0.5$: (*a*) time series, the initial separation in x is 0.001 at time $t = 750$; (*b*) phase projection.

contrast to linear vibration, is a sensitive dependence on initial conditions, and this figure also shows (dashed line) how initially adjacent chaotic trajectories diverge (exponentially on the average) with time. Given a very small difference in the initial conditions, that is, a perturbation of 0.001 in x at time $t = 750$, the mixing nature of the underlying (chaotic) attractor leads to a large difference in the response. Figure 9*b* shows the data for the original time series as a phase projection. Although the response remains bounded, the loss of predictability for increasing time is apparent.

However, unlike truly random systems, the randomness of a chaotic signal is not caused by external influences but is rather an inherent dynamic (deterministic) characteristic. Poincaré sampling can be used effectively to illustrate the underlying structure of a chaotic response. A number of experimental investigations of chaos have been made in a nonlinear vibrations context including the buckled beam,[4] although early work on nonlinear circuits made a significant contribution.[7] Consider the example of a physical analog of the twin-well potential system described by Eq. (9), that is, a small cart rolling on a curved surface.[23] A chaotic attractor based on experimental data is shown in Fig. 10 where

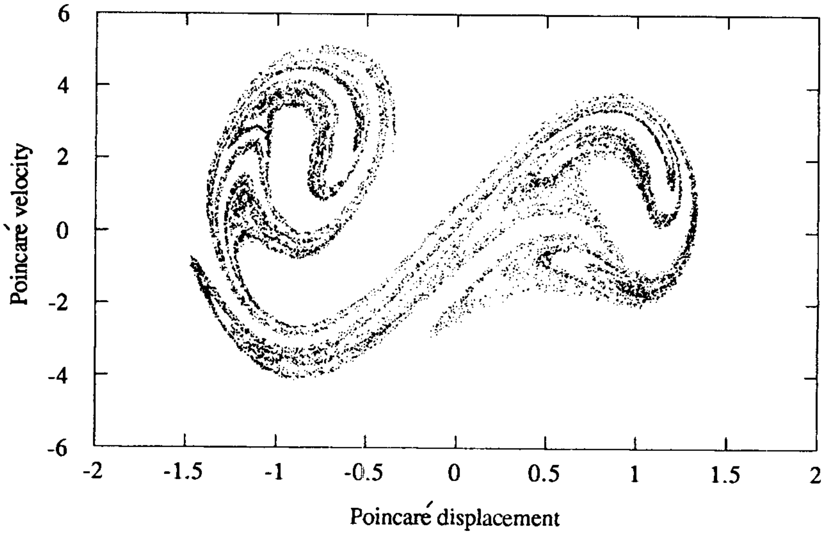


Fig. 10 Poincaré section of the chaotic attractor exhibited by an experimental (mechanical) analogue[23] of Eq. (9).

10,000 Poincaré points (i.e., forcing cycles) are plotted in the phase projection. A similar pattern is obtained by taking a Poincaré section on the data of Fig. 9. The evolution of the trajectory follows a stretching and folding pattern, exhibiting certain fractal characteristics[4], including a noninteger dimension, and can be shown to give a remarkably close correlation with numerical simulation.[23]

For certain parameter ranges catchment regions in forced oscillations also have fractal boundaries, that is, self-similarity at different length scales, although these can occur for periodic attractors. An example is given in Fig. 11 based on numerical integration of Eq. (9) with $b = 0.168$, $\alpha = -0.5$, $\beta = 0.5$, $G = 0.15$, and $\omega = 1$. Here a fine grid of initial conditions is used to determine the basins of attraction for two periodic oscillations, one located within each well. Investigation of all initial conditions is a daunting task, especially for a higher-order system, but certain efficient techniques have been developed.[24] A relatively rare analytical success used to predict the onset of fractal basin boundaries based on homoclinic tangencies is Melnikov theory.[18]

A further remarkable feature of many nonlinear systems is that they often follow a classic, universal route to chaos. Successive period-doubling bifurcations can occur as a system parameter is changed, and these occur at a constant geometric rate, often referred to as Feigenbaum's ratio. This property is exhibited by a large variety of systems including simple difference equations and maps.[5,13,18] Other identified routes to chaos include quasi-periodicity, intermittency and crises, and chaotic transients and homoclinic orbits.[13]

Two major prerequisites for chaos are (i) the system must be nonlinear and (ii) the system must have at least a three-dimensional phase space. A number of diagnostic numerical tools have been developed to characterize chaotic behavior, other than the Poincaré map. The randomlike nature of a chaotic signal is reflected in a broadband power spectrum with all frequencies participating.[25] The widespread availability of FFT software makes power spectral techniques especially attractive in an experimental situation. Also, the divergence of adjacent trajectories can be measured in terms of Lyapunov exponents.[26] The autocorrelation function also has been used to illustrate the increasing loss of correlation with time lag. Various measures of dimension have been developed as well to establish the relation between fractals, chaos, and dynamical behavior.

3.8 Instability

Analogous to the instability of (static) equilibria under the operation of one control parameter, periodic (dynamic) cycles also lose their stability in a small number of generic ways. Figure 12 summarizes the typical stability transitions encountered in nonlinear forced vibration problems. Here, Poincaré sampling is used to obtain information on local transient behavior in terms of char-

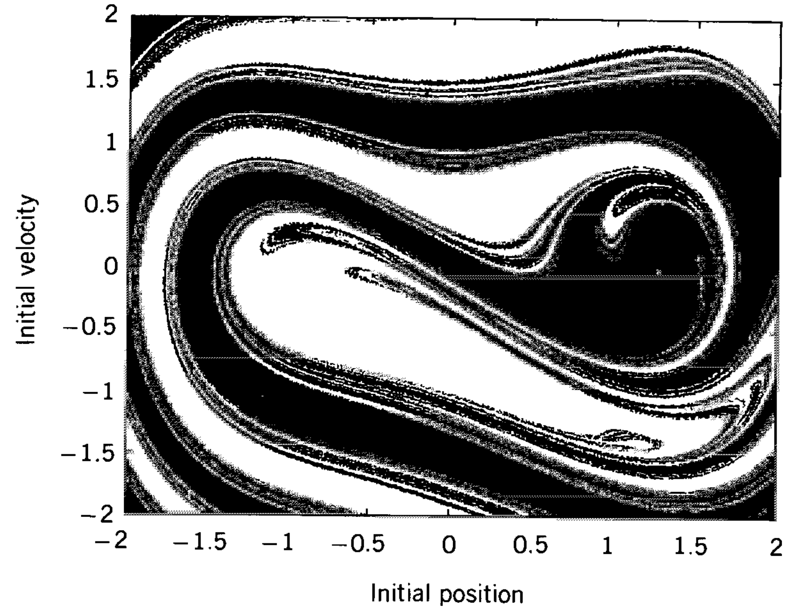


Fig. 11 Fractal basin boundary for Duffing's equation based on numerical simulation ($b = 0.168$, $\alpha = -0.5$, $\beta = 0.5$, $\omega = 1$, $G = 0.15$). The black (white) regions represent initial conditions that generate a transient leading to the periodic attractor in the right (left) hand well.

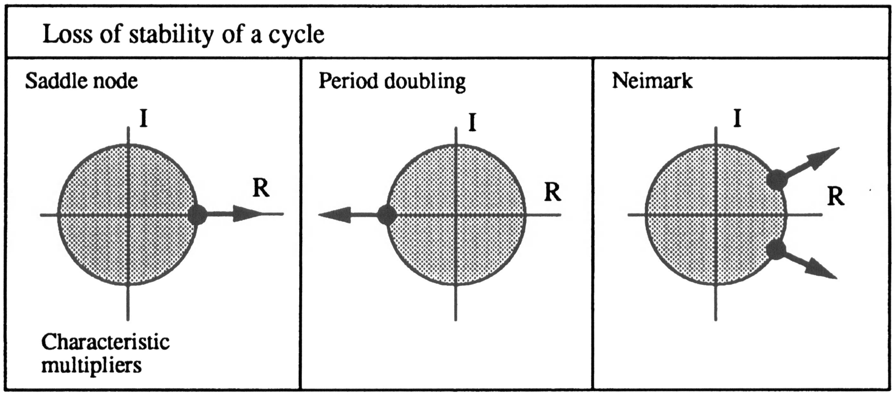


Fig. 12 Generic instabilities of cycles under the operation of a single control.

acteristic multipliers that describe the rate of decay of transients onto a periodic attractor, and hence, in the complex plane characterizes the stability properties of a cycle. This is familiar as the root locus in control theory based on the z-transform.[21] Slowly changing a system parameter may cause penetration of the unit circle and hence the local growth of perturbations. The cyclic fold is the underlying mechanism behind the resonant amplitude jump phenomenon and the flip bifurcation leads to subharmonic motion and may initiate the period-doubling sequence. The Neimark bifurcation is much less commonly encountered in mechanical vibration. Also, other types of instability are possible for nongeneric systems, for example, a perfectly symmetric, periodically forced pendulum may exhibit a symmetry-breaking pitchfork bifurcation.[18]

Again these instability phenomena can also be analyzed using approximate analytical techniques. Small perturbations about a steady-state solution are used to obtain a variational (Mathieu-type) equation, the stability of which can then be determined using Floquet theory or Routh–Hurwitz criteria[1, 14, 16].

4 SUMMARY

Although most of the foregoing examples are drawn from relatively simple single-degree-of-freedom systems, in principle, many of these techniques can be extended to systems of higher order. However, classification, analysis, and computation becomes increasingly difficult, especially for spatially extended systems modeled by partial differential equations. The study of nonlinear problems in vibrations has a solid foundation, especially based on classic analytical techniques,[7, 16, 17] but the scope of chaotic vibration has opened up a whole new range of techniques and applications, and even some potential uses for chaos based on control.[27] However, many problems are still ill-understood: This is even more true for nonlinear systems under random influences.[13, 25]

REFERENCES

1. D. W. Jordan and P. Smith, *Nonlinear Ordinary Differential Equations*, Clarendon Press, Oxford, 1987.
2. R. E. D. Bishop, *Vibration*, Cambridge University Press, Cambridge, 1979.
3. W. T. Thomson, *Theory of Vibration with Applications*, Prentice-Hall, Englewood Cliffs, NJ, 1981.
4. F. C. Moon, *Chaotic and Fractal Dynamics*, Wiley, New York, 1993.
5. J. M. T. Thompson and H. B. Stewart, *Nonlinear Dynamics and Chaos*, Wiley, London 1986.
6. J. L. Singe and B. A. Griffith, *Principles of Mechanics*, McGraw-Hill, New York, 1942.
7. C. Hayashi, *Nonlinear Oscillations in Physical Systems*, McGraw-Hill, New York, 1964.
8. Van der Pol, B., "On Relaxation Oscillations," *Phil. Mag.*, Vol. 7, No. 2, 1926, pp. 978–992.
9. J. W. S. Rayleigh, *The Theory of Sound*, Dover, New York, 1896.
10. J. P. Den Hartog, *Mechanical Vibrations*, McGraw-Hill, New York, 1934.
11. E. H. Dowell, "Flutter of a Buckled Plate as an Example of Chaotic Motion of a Deterministic Autonomous System," *J. Sound Vib.*, Vol. 147, 1982, pp. 1–38.
12. A. B. Pippard, *Response and Stability*, Cambridge University Press, Cambridge, 1985.
13. A. J. Lichtenberg and M. A. Lieberman, *Regular and Chaotic Dynamics*, Springer-Verlag, New York, 1992.
14. A. H. Nayfeh and D. T. Mook, *Nonlinear Oscillations*, Wiley, New York, 1979.
15. C. W. Gear, *Numerical Initial-Value Problems in Ordinary Differential Equations*, Prentice-Hall, Englewood Cliffs, NJ, 1971.

16. N. Minorsky, *Nonlinear Oscillations*, Van Nostrand, Princeton, 1962.
17. J. J. Stoker, *Nonlinear Vibrations*, Interscience, New York, 1950.
18. J. Guckenheimer and P. J. Holmes, *Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields*, Springer-Verlag, New York, 1983.
19. S. W. Shaw and P. J. Holmes, "A Periodically Forced Piece-Wise Linear Oscillator," *J. Sound Vib.*, Vol. 90, 1983, pp. 129–144.
20. P. V. Bayly and L. N. Virgin, "An Experimental Study of an Impacting Pendulum," *J. Sound Vib.*, Vol. 164, 1993, pp. 364–374.
21. A. I. Mees, *Dynamics of Feedback Systems*, Wiley, Chicester, 1981.
22. Y. Ueda, "Steady Motions Exhibited by Duffing's Equation: A Picture Book of Regular and Chaotic Motions," in *New Approaches to Nonlinear Problems in Dynamics*, P. J. Holmes (Ed.), SIAM, Philadelphia, 1980, pp. 311–322.
23. J. A. Gottwald, L. N. Virgin, and E. H. Dowell, "Experimental Mimicry of Duffing's Equation," *J. Sound Vib.*, Vol. 158, 1992, pp. 447–467.
24. C. S. Hsu, *Cell-to-Cell Mapping*, Springer-Verlag, New York, 1987.
25. D. E. Newland, *An Introduction to Random Vibrations and Spectral Analysis*, 2nd ed., Longman, London, 1984.
26. A. Wolf, J. B. Swift, H. L. Swinney, and J. Vastano, "Determining Lyapunov Exponents from a Time Series," *Phys. D*, Vol. 16, 1985, pp. 285–317.
27. E. Ott, C. Grebogi, and J. A. Yorke, "Controlling Chaos," *Phys. Rev. Lett.*, Vol. 64, 1990, pp. 1196–1199.

66

RANDOM VIBRATION

DAVID E. NEWLAND

1 INTRODUCTION

1.1 Random Vibration

Random vibration is an essentially unpredictable phenomenon. Three examples are the response of an offshore structure to excitation from the waves, the response of an aircraft wing to its turbulent boundary layer, and the response of a wheeled vehicle to an uneven road. The assumption that excitation from the waves or a turbulent boundary layer or an uneven road is random is a recognition of our lack of knowledge of the underlying characteristics of this excitation. For example, the surface topography of a given length of road may be measured to any desired accuracy. The excitation from that length of road is then deterministic (not random) because its surface may be precisely described. But it is not practicable to measure the topography of miles of roads precisely, and it is convenient to characterize road surfaces generally by statistical parameters that can be estimated from experimental measurements on particular representative lengths of road. In other cases, for example excitation by sea waves, we may not yet have the physical modeling skills to predict wave loading at any given time. However, measurements can be made that, when taken over a long enough period, allow wave loading to be represented statistically, and accurate predictions then become possible using probabilistic theory.

For constant-parameter, damped, linear systems there is a comprehensive body of theory for calculating how a dynamical system responds to random excitation. Much of this chapter describes the results of this linear theory, including the analysis of peaks and crossings and first-passage and fatigue problems. There are many important results of great practical value.

For time-varying and nonlinear systems, the theory is much less complete and remains a field of active research. Section 8 touches on this by describing three of the principal methods of theoretical analysis for such problems. Where theoretical predictions are impracticable, numerical experiments have become increasingly important. These use pseudorandom excitation generated by so-called Monte Carlo methods. The chapter ends with a brief introduction to this approach.

First it is necessary to describe the statistical structure of a continuously varying random signal.

1.2 Sample Averages

Experimental data on a quantity that is changing randomly, for example the pressure at a point in a turbulent boundary layer, may be represented as a time history in which the magnitude of the quantity concerned $x(t)$ is plotted as a function of time t. As time passes, the character of what is happening may be changing, as for instance when an aircraft descends from high altitude to land. In that case, the time history is from a *nonstationary* random process whose statistical characteristics are changing with time. For a *stationary* random process, the character of the time history does not change with time. Then average values may be obtained by integrating over time so that, for example, the *mean-square* value of $x(t)$ is given by

$$\langle x^2 \rangle = \frac{1}{T} \int_0^T x^2(t)\, dt, \qquad T \to \infty. \tag{1}$$

This is called a *sample average* because it is a result derived from a single time history, or *sample function*,

---

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker
Copyright © 1997 John Wiley & Sons, Inc.

of the random process. Although conceptually straightforward, the sample average formula (1) is only exact when the integration time T is infinite. This raises theoretical difficulties that have been overcome by introducing the idea of an ensemble of sample functions.

1.3 Ensemble Averages

An *ensemble* is a collection of infinitely many sample functions that together make up a *random process*. These separate sample functions are different time histories recorded for the same system: for example, recordings of the boundary layer pressure at the same point on identical aircraft all flying the same route simultaneously. An *ensemble average* is defined as the average value calculated for all the sample functions at a specified time t. The expectation operator E is often used to denote an ensemble average, and $E[x^2(t)]$ is then the ensemble-averaged mean-square value of x at time t.

If a random process is *stationary*, all its ensemble averages are independent of time. If, in addition, the process is *ergodic*, then each sample function has the same statistical characteristics as its ensemble, and then

$$E[x^2(t)] = E[x^2] = \langle x^2 \rangle = \frac{1}{T}\int_0^T x^2(t)\,dt, \qquad T \to \infty. \tag{2}$$

2 PROBABILITY DISTRIBUTIONS AND AVERAGES

2.1 Probability Density Functions

The distribution of values at a given time for all the sample functions or *realizations* of a random process is defined by its *first-order probability density function* $p(x)$. The probability that, for any arbitrarily chosen sample function, the value of $x(t)$ at the chosen time t lies in the band of values x to $x + dx$ is then given by $p(x)\,dx$. Since the probability that x will lie somewhere between $+\infty$ and $-\infty$ is unity (it is a certainty), it follows that

$$\int_{-\infty}^{\infty} p(x)\,dx = 1. \tag{3}$$

This concept can be extended to distributions of more than one random variable. For example, the joint distribution of the values of all the sample functions of a random process at two different times is defined by a *second-order probability density function*. The probability that, for any arbitrarily chosen sample function, the value of $x(t_1) = x_1$ at time t_1 lies in the band of values x_1 to $x_1 + dx_1$ and that, simultaneously, the value of $x(t_2) = x_2$ at time t_2 (for the same sample function) lies in the band x_2 to $x_2 + dx_2$ is given by $p(x_1, x_2)\,dx_1\,dx_2$. Since x_1 and x_2 must both have values between $+\infty$ and $-\infty$, it follows as before that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x_1, x_2)\,dx_1\,dx_2 = 1. \tag{4}$$

Ensemble averages can be calculated directly from the relevant probability density functions, so that, for example,

$$E[x^2] = \int_{-\infty}^{\infty} x^2 p(x)\,dx \tag{5}$$

and

$$E[x_1, x_2] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 p(x_1, x_2)\,dx_1\,dx_2. \tag{6}$$

The second result gives the average value of the product of two values of the same sample function measured at times t_1 and t_2.

2.2 Gaussian Distributions

When a random process results from the combined action of infinitely many independent infinitesimal contributions (e.g., rain on the roof), then according to the *central limit theorem* the probability distributions for the process will be *Gaussian* (or *normal*). Then the probability density functions have the well-known bell shape. The first-order Gaussian probability density function (Fig. 1*a*) is defined by

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-m)^2}{2\sigma^2}\right] \tag{7}$$

and the second-order function (Fig. 1*b*) by

$$p(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{12}^2}} \cdot \exp\left[-\frac{1}{2(1-\rho_{12}^2)}\left\{\frac{(x_1-m_1)^2}{\sigma_1^2} + \frac{(x_2-m_2)^2}{\sigma_2^2} - \frac{2\rho_{12}(x_1-m_1)(x_2-m_2)}{\sigma_1\sigma_2}\right\}\right]. \tag{8}$$

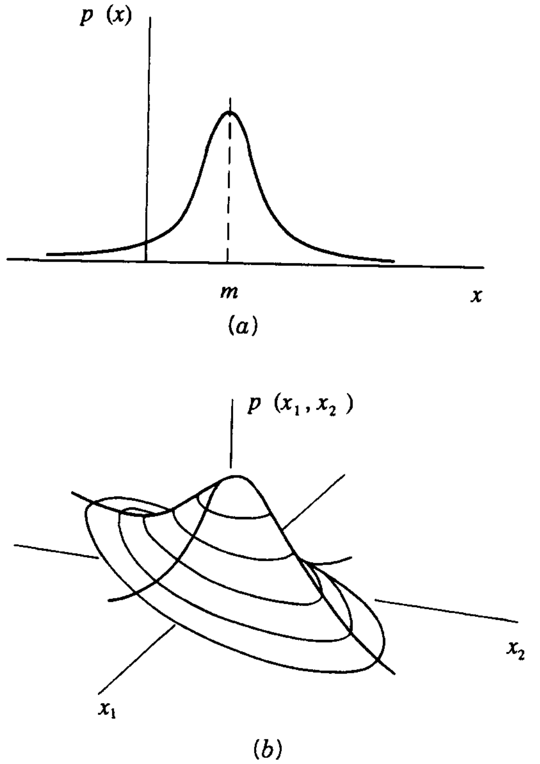


Fig. 1 (*a*) First-order Gaussian probability density curve. (*b*) Second-order Gaussian probability density surface.

In these formulas, the constants are defined as

$$m = E[x], \tag{9}$$

$$\sigma^2 = E[(x - m)^2] = E[x^2] - m^2, \tag{10}$$

and similarly for m_1 and σ_1 when x_1 replaces x and for m_2 and σ_2 when x_2 replaces x. In Eq. (8), the *correlation coefficient* ρ_{12} is defined as

$$\rho_{12} = \frac{E[(x_1 - m_1)(x_2 - m_2)]}{\sigma_1 \sigma_2}. \tag{11}$$

In Eq. (10), the right-hand side may be obtained by noting that

$$E[(x - m)^2] = E[x^2 - 2mx + m^2] = E[x^2] - 2mE[x] + m^2 \tag{12}$$

and using Eq. (9).

2.3 Rayleigh Distribution

Subject to various simplifying assumptions, it can be shown that the distribution of peak values of the sample functions of certain random processes (see Section 6) satisfy a Rayleigh distribution. This describes a random variable whose values lie in the range $0 \le x \le \infty$ (no negative values) and whose probability density function is defined by

$$p(x) = \frac{x}{x_0^2} \exp\left(-\frac{x^2}{2x_0^2}\right), \qquad 0 \le x \le \infty. \tag{13}$$

The shape of this function is shown in Fig. 2*a*. There is a low probability density for very small and very large values of x, signifying that these values are uncommon and that most of the values lie in the region of $x = x_0$.

2.4 Weibull Distribution

Experimental measurements is practical cases (e.g., the vibration of structures under wind loading and the wave-induced bending of ships) have shown that the Weibull distribution may be a better representation of the distribution of peak values of the sample functions of some random processes. The Weibull probability density function is defined by

$$p(x) = k(\ln 2) \frac{x^{k-1}}{X_0^k} \exp\left[-(\ln 2)\left(\frac{x}{X_0}\right)^k\right], \tag{14}$$

where the coefficient k is chosen so that the shape of the resulting probability density curve (Fig. 2*b*) matches the experimental data as closely as possible. In Fig. 2*b*, the curves are plotted against x/X_0 as abscissa, so that $p(x/X_0)$ is drawn instead of $p(x)$, defined by Eq. (14). In order to ensure that the area under the probability density curve remains unity, according to Eq. (3), it is necessary to put

$$p\left(\frac{x}{X_0}\right) = X_0 p(x). \tag{15}$$

When $k = 2$, the Weibull distribution becomes a Rayleigh distribution with the constants X_0 and x_0 related by

$$X_0^2 = 2(\ln 2)x_0^2. \tag{16}$$

As the value of the coefficient k is increased, it can be seen from Fig. 2*b* that most of the values of x lie close to X_0, which can be shown[1] to be the median value of x.

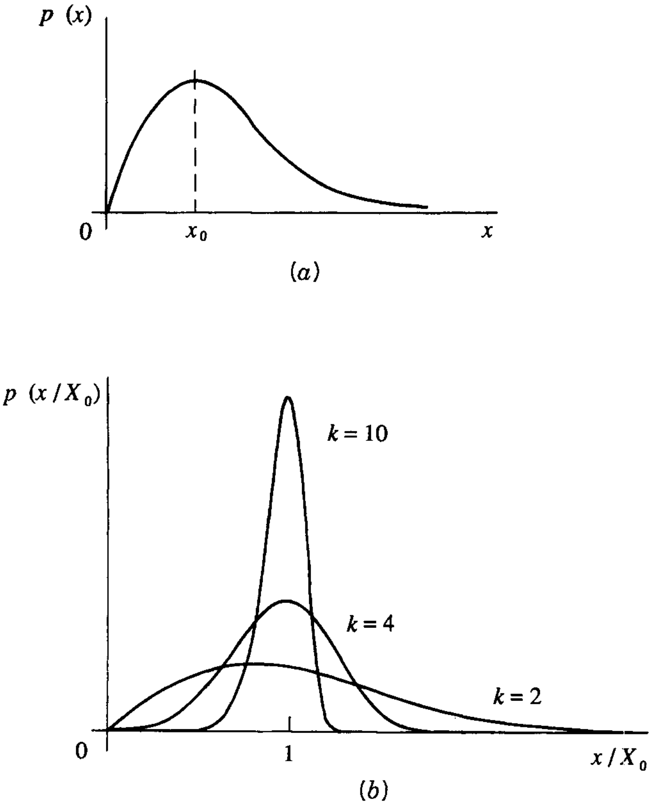


Fig. 2 (*a*) Rayleigh probability density curve. (*b*) Weibull probability density curves.

When $k = \infty$, all the values are X_0. This would be the case for a random process each of whose sample functions consists of a harmonic wave of constant amplitude X_0 but different phases randomly varying from sample function to sample function.

2.5 Distribution Functions

The *probability distribution function* $P(x)$ is defined as the integral of its corresponding density function $p(x)$ by the equation

$$P(x) = \int_{-\infty}^{x} p(\xi)\, d\xi, \tag{17}$$

where ξ is a dummy variable for x. Hence $P(x)$ gives the probability that a random variable lies in the range $-\infty$ to x. For a Gaussian distribution, $P(x)$ does not have a closed form. Values have to be obtained by computation (many vibration analysis software packages include suitable algorithms) or may be obtained from the extensive printed tables of results.[2] For the Weibull distribution

$$P(x) = 1 - \exp\left[-(\ln 2)\left(\frac{x}{X_0}\right)^k\right]. \tag{18}$$

This result may easily be checked against Eq. (14) by using Eq. (17), which, after differentiation, becomes

$$\frac{dP(x)}{dx} = p(x). \tag{19}$$

Gaussian and Weibull distribution functions are illustrated in Figs. 3*a* and 3*b*.

2.6 Conditional Probability

Calculations on the statistics of random vibration response use the notion of *conditional probability*. This describes the probability of one event occurring on the condition that a second event also occurs. For example, suppose that a sample function $x(t)$ from a random process is examined at times t_1 and t_2. We may want to know what the probability is that $x(t_1)$ lies in the band of values x_1 to $x_1 + dx_1$ on the condition that $x(t_2)$ lies in the band x_2 to $x_2 + dx_2$. Many of the sample functions will not satisfy either of these conditions. We are interested only in those samples that satisfy both the conditions.

From the ensemble of all sample functions, we separate out a subensemble of sample functions that satisfy one of the two conditions, for example $x_2 < x(t_2) < x_2 + dx_2$. Theoretically, since we have an infinite number of sample functions in the original ensemble, this subensemble will also have an infinite number of sample functions. The *conditional probability density* function, which is represented as $p(x(t_1)|x(t_2))$, then describes the distribution of the random variable $x(t_1)$ within the subensemble for which all the $x(t_2)$ lie in the band x_2 to $x_2 + dx_2$.

The probability that a sample chosen at random from the complete ensemble lies in the subensemble is $p(x_2)\, dx_2$, where $p(x_2)$ means the probability density function for the distribution of $x(t)$ at $t = t_2$. The probability that a sample from this subensemble lies in the band x_1 to $x_1 + dx_1$ is $p(x_1|x_2)\, dx_1$, where $p(x_1|x_2)$ means the probability density for the distribution of $x(t)$ at $t = t_1$ for samples in the subensemble. But the probability that a sample from the complete ensemble lies in both bands is $p(x_1, x_2)\, dx_1\, dx_2$, where $p(x_1, x_2)$ is the second-order probability density function for the joint distribution of $x(t)$ at t_1 and t_2. Hence it follows[1] that

$$p(x_1|x_2)p(x_2) = p(x_1, x_2). \tag{20}$$

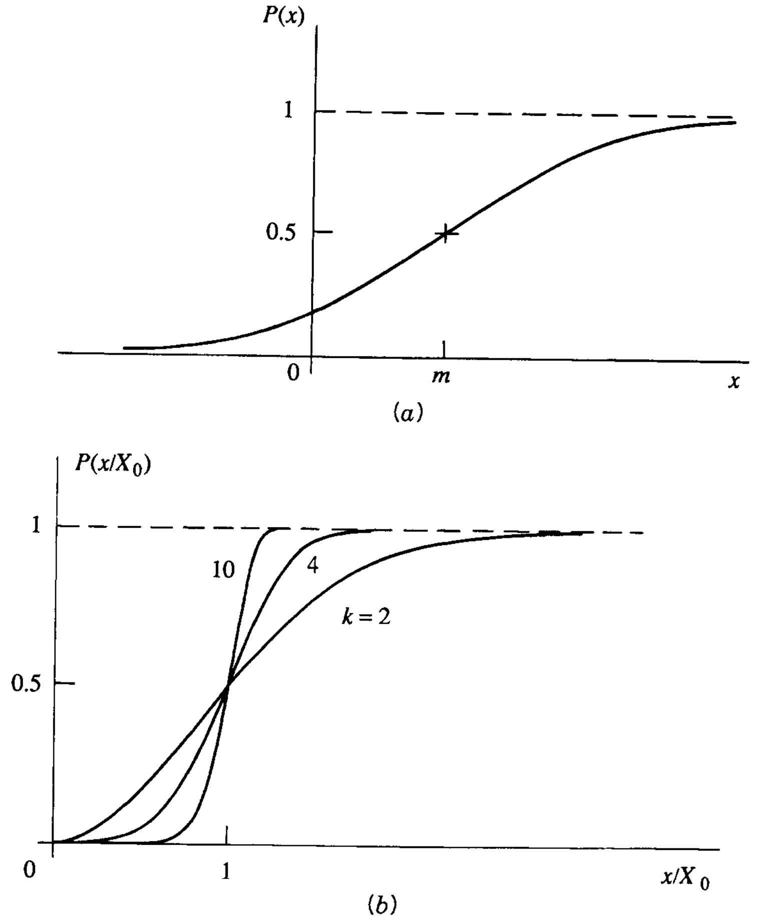


Fig. 3 (*a*) Gaussian probability distribution function. (*b*) Weibull probability distribution functions.

If the conditional probability density function is independent of its condition, so that

$$p(x_1 | x_2) = p(x_1), \tag{21}$$

then it follows that

$$p(x_1, x_2) = p(x_1)p(x_2). \tag{22}$$

When a second-order probability density function can be factored into two first-order probability density functions, as in Eq. (22), the two random variables involved are said to be *statistically independent* of each other.

3 CORRELATION FUNCTIONS

3.1 Autocorrelation

The starting point for studying the characteristics of the sample functions that make up a random process is to examine how they change with time (or with whatever is the independent base parameter). This is done by using the *autocorrelation function*, which is the ensemble average of the product of values of each sample function at two different times, for example, $E[x(t_1)x(t_2)]$. This average is usually denoted by $R_{xx}(t_1, t_2)$, where the double subscript denotes the quantity being correlated and the arguments t_1, t_2 denote the times at which $x(t)$ has been sampled.

Using Eq. (6) gives

$$R_{xx}(t_1, t_2) = E[x(t_1)x(t_2)]$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 p(x_1, x_2)\, dx_1\, dx_2, \tag{23}$$

where x_1 means $x(t_1)$ and x_2 means $x(t_2)$.

For a stationary process, the autocorrelation function must be independent of absolute time, so that if

$$\tau = t_2 - t_1, \tag{24}$$

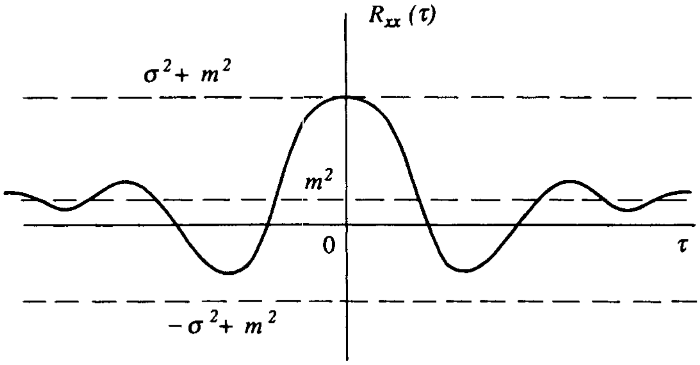


Fig. 4 Properties of the autocorrelation function $R_{xx}(\tau)$ for a stationary random process $x(t)$.

then $R_{xx}(t_1, t_2)$ becomes $R_{xx}(\tau)$ and is a function only of the separation between the two measuring points t_1 and t_2.

It can be shown[1] that, for a stationary process, $R_{xx}(\tau)$ has the important properties illustrated in Fig. 4. Its upper bound is its value at $\tau = 0$, which, from Eq. (23), is

$$E[x^2(t_1)] = E[x^2] = \sigma^2 + m^2 \tag{25}$$

from Eq. (10). Its lower bound is $-\sigma^2 + m^2$. When $|\tau| \rightarrow \infty$, $R_{xx}(\tau) \rightarrow m^2$. And since $R_{xx}(\tau)$ depends only on the magnitude of the time separation, it follows that $R_{xx}(\tau) = R_{xx}(-\tau)$, so that $R_{xx}(\tau)$ is an even function of τ.

3.2 Cross-correlation

When there are two corresponding random processes, for example $x(t)$ and $y(t)$, in which each sample function of $x(t)$ is associated with a corresponding sample function of $y(t)$, then the *cross-correlation functions* for the two processes are defined as

$$R_{xy}(t_1, t_2) = E[x(t_1)y(t_2)] \tag{26}$$

and

$$R_{yx}(t_1, t_2) = E[y(t_1)x(t_2)] \tag{27}$$

If both processes $x(t)$ and $y(t)$ are stationary, then the cross-correlation functions depend only on τ [Eq. (24)], and

$$R_{xy}(\tau) = E[x(t)y(t + \tau)] = R_{yx}(-\tau)], \tag{28}$$

$$R_{yx}(\tau) = E[y(t)x(t + \tau)] = R_{xy}(-\tau)]. \tag{29}$$

It can be shown[1] that the upper and lower bounds of $R_{xy}(\tau)$ and $R_{yx}(\tau)$ are $\sigma_x\sigma_y + m_x m_y$ and $-\sigma_x\sigma_y + m_x m_y$, respectively, where m_x, m_y are the means and σ_x, σ_y the standard deviations of x and y according to Eqs. (9) and (10). The asymptotic value to which the cross-correlation functions converge as $|\tau| \rightarrow \infty$ is $m_x m_y$.

These properties are illustrated in Fig. 5.

The main importance of correlation functions is that they provide the theoretical basis for spectral analysis (Section 4).

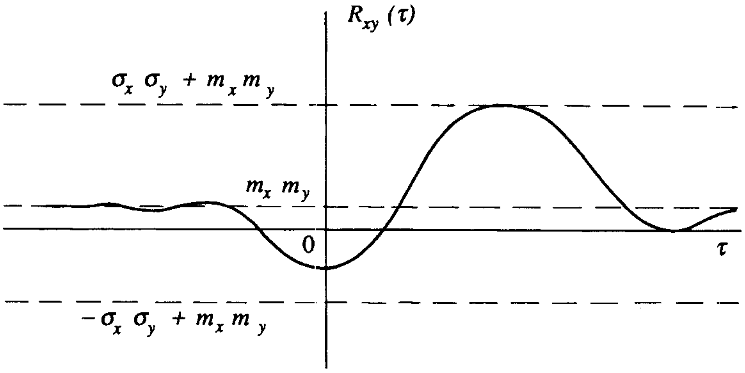


Fig. 5 Properties of the cross-correlation function $R_{xy}(\tau)$ for two stationary random processes $x(t)$ and $y(t)$.

4 SPECTRAL ANALYSIS

4.1 Spectral Density

The (auto) *spectral density* of a stationary random process $x(t)$ is defined as the Fourier transform of its autocorrelation function, so that

$$S_{xx}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} R_{xx}(\tau)e^{-i\omega\tau}\, d\tau. \tag{30}$$

The inverse Fourier transform allows the autocorrelation function to be recovered from the spectral density, according to

$$R_{xx}(\tau) = \int_{-\infty}^{\infty} S_{xx}(\omega)e^{i\omega\tau}\, d\omega. \tag{31}$$

The spectral density $S_{xx}(\omega)$ is a function of angular frequency ω. It can be shown[1] that it is a real, even, and nonnegative function of ω and therefore has the form illustrated in Fig. 6. The term $S_{xx}(\omega)$ is often called the *mean-square spectral density* because, by putting $\tau = 0$ in Eq. (31),

$$E[x^2] = R_{xx}(\tau = 0) = \int_{-\infty}^{\infty} S_{xx}(\omega)\, d\omega, \tag{32}$$

so that the shaded area in Fig. 6 is equal to the magnitude of the mean-square value of $x(t)$. The name *power spectral density* (or PSD) is also used and derives from electrical applications, where mean square was a measure of power but, strictly, this is a misnomer in the general case.

4.2 Cross-spectral Density

The cross-spectral densities $S_{xy}(\omega)$ and $S_{yx}(\omega)$ are defined as the Fourier transforms of their corresponding cross-correlation functions $R_{xy}(\tau)$ and $R_{yx}(\tau)$, so that

$$S_{xy}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} R_{xy}(\tau)e^{-i\omega\tau}\, d\tau, \tag{33}$$

$$S_{yx}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} R_{yx}(\tau)e^{-i\omega\tau}\, d\tau. \tag{34}$$

On account of Eqs. (28) and (29), it follows from these definitions that

$$S_{xy}(\omega) = S^*_{yx}(\omega) \quad \text{and} \quad S_{yx}(\omega) = S^*_{xy}(\omega), \tag{35}$$

where the asterisk denotes complex conjugate. The cross-correlation functions may be regained from $S_{xy}(\omega)$ and $S_{yx}(\omega)$ by the inverse transform relations

$$R_{xy}(\tau) = \int_{-\infty}^{\infty} S_{xy}(\omega)e^{i\omega\tau}\, d\omega, \tag{36}$$

$$R_{yx}(\tau) = \int_{-\infty}^{\infty} S_{yx}(\omega)e^{i\omega\tau}\, d\omega. \tag{37}$$

Usually cross-spectral densities are complex (have real and imaginary parts) and they do not have the simple properties of (auto) spectral density functions.

5 RESPONSE OF LINEAR SYSTEMS

5.1 Frequency Response Functions

Consider a time-invariant linear dynamical system with a typical input $x_s(t)$, for example an applied load or voltage, and a typical output $y_r(t)$, for example a displacement or an electrical current. If only input $x_s(t)$ is excited with a unit harmonic excitation, then using phasor nota-

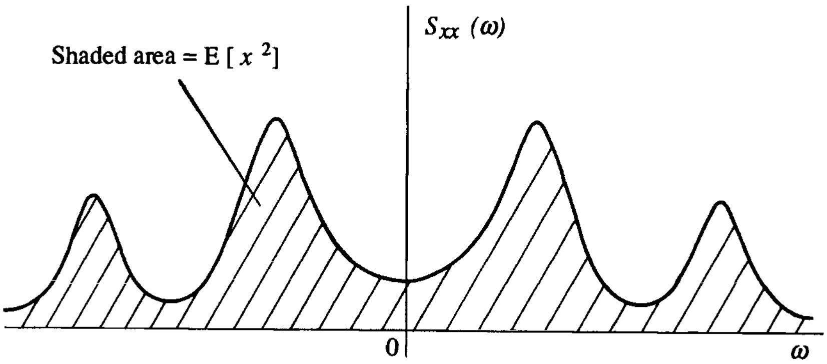


Fig. 6 Properties of the spectral density function.

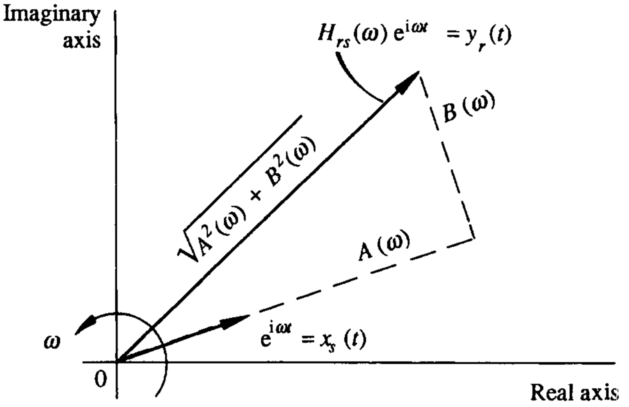


Fig. 7 Phasor representation of the frequency response function $H_{rs}(\omega)$ of a time-invariant, passive, linear system.

tion,

$$x_s(t) = e^{i\omega t}. \tag{38}$$

After steady conditions have been reached (the system is assumed to be stable so that transients always die away eventually), the response $y_r(t)$ can be represented by

$$y_r(t) = H_{rs}(\omega)e^{i\omega t}, \tag{39}$$

where $H_{rs}(\omega)$ is the *frequency response function* that describes the connection between input s and output r. It is a complex function of frequency that relates the amplitude and phase of output r to the assumed unit amplitude and arbitrary phase of input s. This relationship is illustrated in the phasor diagram of Fig. 7. The input $x_s(t)$ is represented as a phasor of unit length rotating counterclockwise at angular speed ω. Then, assuming that

$$H_{rs}(\omega) = A(\omega) + iB(\omega), \tag{40}$$

where $A(\omega)$ and $B(\omega)$ are real and positive, the output $y_r(t)$ is a rotating phasor of length $\sqrt{A^2(\omega) + B^2(\omega)}$ rotating in the same direction at speed ω and leading the excitation by the angle $\tan^{-1}[B(\omega)/A(\omega)]$.

The time dependence of $x_s(t)$ and $y_r(t)$ is given by the projection of the rotating phasors onto either the real (horizontal) axis or the imaginary (vertical) axis in Fig. 7. The selection of which one is arbitrary.

If there are n inputs and m outputs, then there are $n \times m$ frequency response functions that together make up the *frequency response matrix*[3] $\mathbf{H}(\omega)$.

5.2 Impulse Response Functions

When input $x_s(t)$ is given a unit impulse of excitation at $t = 0$ so that

$$x_s(t) = \delta(t), \tag{41}$$

with all the other inputs zero and with the system initially settled, the response $y_r(t)$ is represented by

$$y_r(t) = h_{rs}(t). \tag{42}$$

This real function of time $h_{rs}(t)$ is called the *impulse response function* of output r to input s. It is zero for $t < 0$, and since the linear system is assumed to be passive, it decays to zero as $t \to \infty$.

When there are n inputs and m outputs, there are $n \times m$ impulse response functions that together make up the *impulse response matrix*[3] $\mathbf{h}(t)$.

5.3 Fourier Transform Relationship

It can be shown[3] that each frequency response function is the Fourier transform of its corresponding impulse response function, so that

$$\mathbf{H}(\omega) = \int_{-\infty}^{\infty} \mathbf{h}(t)e^{-i\omega t}\,dt \tag{43}$$

and, conversely, that

$$\mathbf{h}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbf{H}(\omega)e^{i\omega t}\,d\omega. \tag{44}$$

In this case, the factor $1/2\pi$ appears in the inverse transform equation rather than in the transform equation as, for example, in Eq. (30).

Both $\mathbf{H}(\omega)$ and $\mathbf{h}(t)$, and the relationship between them, are needed in establishing how linear systems respond to random excitation.

5.4 Response of a SISO System

The simplest case is of a *single-input, single-output* (SISO) time-invariant passive linear system. This is assumed to be excited by stationary random excitation $x(t)$ with mean level $E[x]$, autocorrelation $R_{xx}(\tau)$, and spectral density $S_{xx}(\omega)$.

It can be shown[1] that the corresponding statistics of the response are

$$E[y] = H(\omega = 0)E[x], \tag{45}$$

$$R_{yy}(\tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(t_1)h(t_2)R_{xx}(\tau + t_1 - t_2) \cdot dt_1\,dt_2, \tag{46}$$

$$S_{yy}(\omega) = |H(\omega)|^2 S_{xx}(\omega), \tag{47}$$

where $H(\omega = 0)$ is the frequency response function evaluated at $\omega = 0$, $|H(\omega)|$ is the modulus of the frequency response function, and $h(t_1)$, $h(t_2)$ are the impulse response functions evaluated at two different times t_1 and t_2.

For comparison with the *multiple-input, multiple-output* (MIMO) case, it is convenient to write Eq. (45) as

$$m_y = H(\omega = 0)m_x. \tag{48}$$

5.5 Response of a MIMO System

When there are n inputs and m outputs, the corresponding results are[1]

$$m_{y_r} = \sum_{s=1}^{n} H_{rs}(\omega = 0)m_{x_s}, \tag{49}$$

$$R_{y_q y_r} = \sum_{s=1}^{n} \sum_{t=1}^{n} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdot\, h_{qs}(t_1)h_{rt}(t_2)R_{x_s x_t}(\tau + t_1 - t_2)\, dt_1\, dt_2, \tag{50}$$

$$S_{y_q y_r}(\omega) = \sum_{s=1}^{n} \sum_{t=1}^{n} H_{qs}^{*}(\omega)H_{rt}(\omega)S_{x_s x_t}(\omega), \tag{51}$$

where now $R_{y_q y_r}$ is the cross-correlation function and $S_{y_q y_r}$ is the cross-spectral density for the two outputs $y_q(t)$ and $y_r(t)$.

These results can be expressed in matrix notation by defining the following:

$\mathbf{m}_x = n \times 1$	vector of the mean values of the n inputs
$\mathbf{m}_y = m \times 1$	vector of the mean values of the m outputs
$\mathbf{h}(t) = m \times n$	matrix of the $m \times n$ impulse response functions relating m outputs to the n inputs
$\mathbf{H}(\omega) = m \times n$	matrix of the $m \times n$ frequency response functions relating m outputs to the n inputs
$\mathbf{R}_{xx}(\tau) = n \times n$	matrix of the $n \times n$ auto- and cross-correlation functions for the inputs
$\mathbf{R}_{yy}(\tau) = m \times m$	matrix of the $m \times m$ auto- and cross-correlation functions for the outputs
$\mathbf{S}_{xx}(\omega) = n \times n$	matrix for the $n \times n$ auto- and cross-spectral density functions for the inputs
$\mathbf{S}_{yy}(\omega) = m \times m$	matrix of the $m \times m$ auto- and cross-spectral density functions for the outputs

The input–output results then become

$$\mathbf{m}_y = \mathbf{H}(\omega = 0)\mathbf{m}_X, \tag{52}$$

$$\mathbf{R}_{yy}(\tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{h}(t_1)\mathbf{R}_{xx}(\tau + t_1 - t_2) \cdot \mathbf{h}^{\mathrm{T}}(t_2)\, dt_1\, dt_2, \tag{53}$$

$$\mathbf{S}_{yy}(\omega) = \mathbf{H}^{*}(\omega)\mathbf{S}_{xx}(\omega)\mathbf{H}^{\mathrm{T}}(\omega). \tag{54}$$

In addition to these formulas for the output statistics, it is also possible to calculate the correlation that there is between a typical output $y_r(t)$ and one of the inputs $x_s(t)$. This leads to the following results[1]:

$$\mathbf{R}_{yx}(\tau) = \int_{-\infty}^{\infty} \mathbf{h}(t)\mathbf{R}_{xx}(\tau + t)\, dt \tag{55}$$

$$\mathbf{S}_{yx}(\omega) = \mathbf{H}^{*}(\omega)\mathbf{S}_{xx}(\omega). \tag{56}$$

where $\mathbf{R}_{yx}(\tau)$ is the $m \times n$ matrix of the $m \times n$ cross-correlation functions between the m outputs $y_r(t)$, $r = 1, \ldots, m$, and the n inputs $x_s(t)$, $s = 1, \ldots, n$, and $\mathbf{S}_{yx}(\omega)$ is the $m \times n$ matrix of the corresponding $m \times n$ cross-spectral density functions between the outputs and inputs.

Because of the simplicity of Eqs. (47), (54), and (56), calculations in the frequency domain offer the main computational approach for studying many practical problems. A great deal of research has therefore been done on digital spectral processing,[1,4,5] which is now the major tool for practical random vibration calculations.

5.6 Coherence Calculations

If there are unknown sources of noise affecting a system's response, or if the system is nonlinear, the measured spectral densities of the output will differ from those calculated by using Eq. (54). Consider only one output, so that $m = 1$ in Eq. (54) and $S_{yy}(\omega)$ has only one term. If the measured output spectral density is $S_{yy}(\omega)$ and the calculated value is $\tilde{S}_{yy}(\omega)$, then their ratio

$$\frac{\tilde{S}_{yy}(\omega)}{S_{yy}(\omega)} = \eta_{yx}^{2}(\omega) \tag{57}$$

is called the *multiple coherence function* between the output process $y(t)$ and the input processes $x_s(t)$, $s = 1, \ldots, n$. It can be shown[6] that

$$0 \le \eta^2_{yx}(\omega) \le 1 \tag{58}$$

and that $\eta^2_{yx}(\omega)$ is only unity if the system is time invariant, linear, and noise free.

On substituting for the calculated spectral density from Eqs. (54) into (57),

$$\eta^2_{yx}(\omega) = \frac{\mathbf{H}^*(\omega)\mathbf{S}_{xx}(\omega)\mathbf{H}^{\mathrm{T}}(\omega)}{S_{yy}(\omega)}, \tag{59}$$

and substituting for the frequency response matrix $\mathbf{H}(\omega)$ from Eq. (56) give

$$\eta^2_{yx}(\omega) = \frac{\mathbf{S}_{yx}(\omega)\mathbf{S}^{-1}_{xx}(\omega)\mathbf{S}^{\mathrm{T}}_{xy}(\omega)}{S_{yy}(\omega)}. \tag{60}$$

If there is only one input, so that all the matrices have only one term, Eq. (60) becomes

$$\eta^2_{yx}(\omega) = \frac{S_{yx}(\omega)S_{xy}(\omega)}{S_{xx}(\omega)S_{yy}(\omega)}. \tag{61}$$

This gives the single or *ordinary coherence function* between output $y(t)$ and input $x(t)$ at frequency ω. If the system is linear and time invariant and there is only one input, then if the spectra are measured accurately, η^2 will be exactly unity.

Because of the need for accurate measurements, there are practical problems when computing coherence functions by digital analysis.[1,4]

5.7 Probability Distributions

When a time-invariant passive linear system is subjected to Gaussian excitation, it can be shown[7] that the system's response will also be Gaussian. For MIMO systems, the separate inputs have to be jointly Gaussian and the outputs are then also jointly Gaussian (as well as first-order distributions for the individual signals being Gaussian, higher order distributions between two or more signals are also Gaussian).

For non-Gaussian inputs, the output probability distributions can be expressed in terms of a series expansion involving the input probabilities and the system's response characteristics,[8] but the calculations are extremely complex. There are no general closed-form solutions.

The upshot is that, although the spectral density of the output can always be found, calculations on the amplitude distribution of the output may not be practicable unless the excitation is Gaussian. Fortunately, because of the central limit theorem, many naturally occurring processes are Gaussian to a good approximation.

The probability distributions for a Gaussian process are defined if the mean, mean-square, and appropriate correlation functions are known [see Eqs. (7) and (8)]. According to linear system theory, these statistics can always be calculated from the formulas in Section 5.5.

5.8 Continuous Systems

There are many applications where a continuous, stable, time-invariant, linear system, for example a mechanical structure, is subjected to spatially distributed time-stationary random excitation, for example a fluctuating pressure field. Figure 8 shows a notional continuous system subjected to distributed excitation (e.g., pressure) $p(\mathbf{s}, t)$ at position $\mathbf{s}$ and time t and that has a distributed response (e.g., displacement) $y(\mathbf{r}, t)$ at position $\mathbf{r}$ and time t. Corresponding to Eqs. (49)–(51), the response of this system is given by[1]

$$m_y(\mathbf{r}) = \int_R ds\, H(\mathbf{r}, \mathbf{s}, \omega = 0)m_p(\mathbf{s}), \tag{62}$$

$$R_{yy}(\mathbf{r}_1, \mathbf{r}_2, \tau) = \int_R d\mathbf{s}_1 \int_R d\mathbf{s}_2 \int_{-\infty}^{\infty} dt_1 \int_{-\infty}^{\infty} \cdot\, dt_2\, h(\mathbf{r}_1, \mathbf{s}_1, t_1)h(\mathbf{r}_2, \mathbf{s}_2, t_2) \cdot R_{pp}(\mathbf{s}_1, \mathbf{s}_2, \tau + t_1 - t_2), \tag{63}$$

and

$$S_{yy}(\mathbf{r}_1, \mathbf{r}_2, \omega) = \int_R d\mathbf{s}_1 \int_R d\mathbf{s}_2\, H^*(\mathbf{r}_1, \mathbf{s}_1, \omega) \cdot H(\mathbf{r}_2, \mathbf{s}_2, \omega)S_{pp}(\mathbf{s}_1, \mathbf{s}_2, \omega). \tag{64}$$

The finite summations in Eqs. (48)–(51) have been replaced by continuous integrals over the region involved, R. The definitions of terms correspond to those in Section 5.5, where now, $R_{yy}(\mathbf{r}_1, \mathbf{r}_2, \tau)$ denotes the cross-correlation between $y(\mathbf{r}_1, t)$ and $y(\mathbf{r}_2, t + \tau)$ and replaces $R_{y_q y_r}(\tau)$, and similarly for the other terms.

For the above results the excitation is assumed to be stationary in time. It may also be stationary in space (or *homogeneous* in space). In that case the correlation of $p(\mathbf{s}, t)$ at two different positions $\mathbf{s}_1$ and $\mathbf{s}_2$ depends only on the vector difference between these positions $(\mathbf{s}_2 - \mathbf{s}_1)$. The cross-spectral density of the input can then be written as[1]

$$\int_\infty d\gamma\, S_{pp}(\gamma, \omega) \exp i\{\gamma \cdot (\mathbf{s}_2 - \mathbf{s}_1)\}, \tag{65}$$

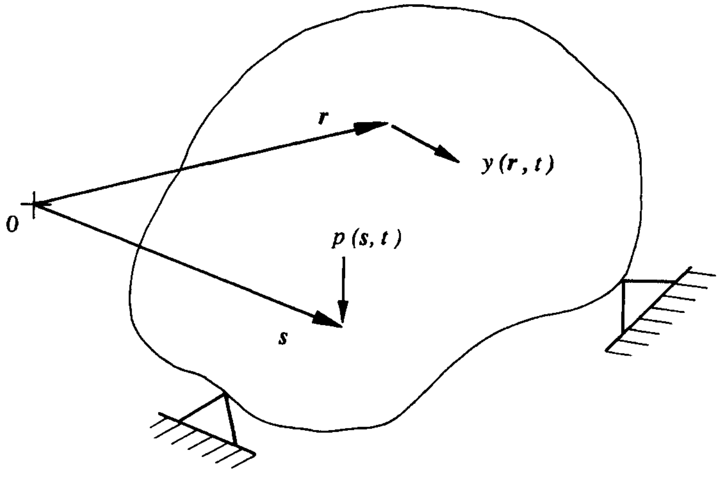


Fig. 8 Continuous dynamic system subjected to distributed excitation $p(s,t)$ in a specified reference direction and having distributed response $y(r,t)$, also in a specified reference direction.

where $S_{pp}(\gamma,\omega)$ is the multidimensional spectral density of the distributed excitation process $p(\mathbf{s},t)$ in terms of the (vector) wavenumber γ and the angular frequency ω.

Substituting Eq. (65) into Eq. (64) leads to the result that

$$S_{yy}(\mathbf{r}_1,\mathbf{r}_2,\omega)=\int_\infty d\gamma\; S_{pp}(\gamma,\omega)G^*(\mathbf{r}_1,\gamma,\omega)G(\mathbf{r}_2,\gamma,\omega), \tag{66}$$

where

$$G(\mathbf{r},\gamma,\omega)=\int_R H(\mathbf{r},\mathbf{s},\omega)\exp\, i(\gamma\cdot\mathbf{s})\,ds. \tag{67}$$

This function $G(\mathbf{r},\gamma,\omega)$ is called the *sensitivity function* for the system. It gives the sensitivity of the system measured at position **r** to distributed excitation that extends over the whole system and is harmonic in space and time at wave vector γ and time frequency ω.

5.9 Normal-Mode Analysis

Normal-mode analysis depends on the experimental result that, in many practical vibration problems, the response of interest is confined mainly to a limited number of natural modes (see Chapter 63). This provides an approximate method of calculating $H(\mathbf{r},\mathbf{s},\omega)$.

If $\mathbf{y}(\mathbf{r},t)$ is a column vector of the geometric components of the response of a continuous system at position **r** and time t and if $\mathbf{U}_j(\mathbf{r})$ is a column vector of the corresponding components of the normal modes, then the normal-mode expansion of the response is given by

$$y(\mathbf{r},t)=\sum_{j=1}^{\infty} q_j(t)\mathbf{U}_j(\mathbf{r}), \tag{68}$$

where $q_j(t)$, $j=1,\ldots,\infty$, are the normal coordinates. The system is assumed to have no energy transfer across the boundaries and classical damping so that a full set of real, orthogonal, normal-mode functions exists. In that case the general vector frequency response function $\mathbf{H}(\mathbf{r},\mathbf{s},\omega)$ can be written in the expansion

$$\mathbf{H}(\mathbf{r},\mathbf{s},\omega)=\sum_{j=1}^{\infty} H_j(\omega)\mathbf{U}_j(\mathbf{r})\mathbf{U}_j^{\mathrm{T}}(\mathbf{s})\mathbf{e}_s, \tag{69}$$

where $H_j(\omega)$ is the modal frequency response function and $\mathbf{e}_s$ is a unit vector identifying the direction of the unit harmonic excitation applied at position **s**.[3] The components of $\mathbf{H}(\mathbf{r},\mathbf{s},\omega)$ are the frequency responses in the three coordinate directions at **r** for unit harmonic excitation applied at **s** in direction $\mathbf{e}_s$.

For classical damping[3] the normal-mode functions are the same as those for the undamped problem and may, in principle, be obtained by solving for the undamped free vibration of the system. In practice, this may prove impossible, and the usual approach is then to establish an approximate model with a finite number of degrees of freedom, usually using the finite-element method, and to solve this new problem exactly.

6 NARROW-BAND PROCESSES

6.1 Properties of Narrow-Band Random Processes

When a strongly resonant system responds to broadband random excitation, its output spectral density falls mainly in a narrow band of frequencies close to the resonant frequency. Since this output is derived by filtering a broadband process, many nearly independent events contribute to it. Therefore, on account of the central limit theorem, the probability distribution of a narrow-band process approaches that of a Gaussian distribution even if the excitation is not Gaussian.

Assuming that the narrow-band process is Gaussian and stationary, then if its spectral density $S_{yy}(\omega)$ is specified, all its other statistics can be derived from $S_{yy}(\omega)$. The corresponding autocorrelation function $R_{yy}(\tau)$ can be calculated by using the inverse Fourier transform relation (31) and the variance σ_y^2 and mean m_y found from the properties of $R_{yy}(\tau)$ illustrated in Fig. 4. Also the statistics of the $\dot{y}(t)$ process can be derived by using the result that, for any stationary process,[1]

$$R_{\dot{y}\dot{y}}(\tau) = -\frac{d^2}{d\tau^2} R_{yy}(\tau). \tag{70}$$

Hence $\sigma_{\dot{y}}^2$ and $m_{\dot{y}}$ can be found.

For any stationary random process it can be shown[1] that $y(t)$ and $\dot{y}(t)$ are uncorrelated, so that, from Eq. (22),

$$p(y, \dot{y}) = p(y)p(\dot{y}). \tag{71}$$

The first-order Gaussian distributions are known when σ and m are specified, so that $p(y)$ and $p(\dot{y})$ are known and, from Eq. (71), so is the second-order Gaussian distribution $p(y, \dot{y})$.

6.2 Crossing Analysis

Figure 9 shows a sample function from a stationary, narrow-band random process. The ensemble average number of up-crossings (crossings with positive slope) of the level $y = a$ in time T, denoted by $N^+(a, T)$, will be proportional to T, so that

$$N^+(a, T) = \nu^+(a)T, \tag{72}$$

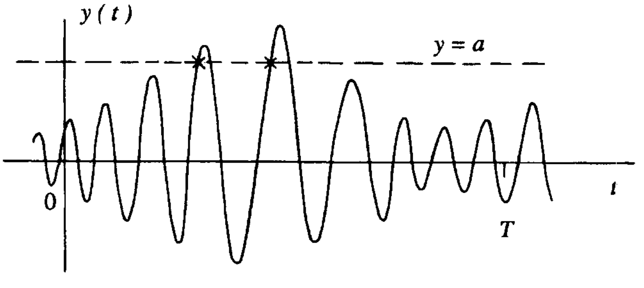


Fig. 9 Up-crossings of a sample function from a narrow-band process $y(t)$.

where $\nu^+(a)$ is the frequency of up-crossings of level $y = a$. It can be shown[1] that

$$\nu^+(a) = \int_0^\infty p(a, \dot{y})\dot{y}\, d\dot{y}, \tag{73}$$

where $p(a, \dot{y})$ is the joint probability density function for the distribution of y and $\dot{y}$ evaluated at $y = a$. By using Eq. (71) and for the Gaussian probability density function (7) with zero mean, this leads to the result that

$$\nu^+(a) = \frac{1}{2\pi} \frac{\sigma_{\dot{y}}}{\sigma_y} \exp\left(-\frac{a^2}{2\sigma_y^2} \right), \tag{74}$$

where σ_y^2 is the variance of $y(t)$ and $\sigma_{\dot{y}}^2$ is the variance of $\dot{y}(t)$.

For a process with zero mean, from Eqs. (10) and (32),

$$\sigma_y^2 = \int_{-\infty}^{\infty} S_{yy}(\omega)\, d\omega. \tag{75}$$

Also it can be shown from Eq. (70) that[1]

$$S_{\dot{y}\dot{y}}(\omega) = \omega^2 S_{yy}(\omega), \tag{76}$$

so that

$$\sigma_{\dot{y}}^2 = \int_{-\infty}^{\infty} \omega^2 S_{yy}(\omega)\, d\omega. \tag{77}$$

Using these results, the average frequency for the $y(t)$ process, which is obtained by putting $a = 0$ in Eq. (74), can be written as

$$\nu^+(0) = \frac{1}{2\pi} \left[\frac{\int_{-\infty}^{\infty} \omega^2 S_{yy}(\omega)\, d\omega}{\int_{-\infty}^{\infty} S_{yy}(\omega)\, d\omega} \right]^{1/2}. \tag{78}$$

This is the ensemble average frequency for the $y(t)$ process. It is only the same as the average frequency along the time axis if the process is ergodic.

6.3 Distribution of Peaks

For a narrow-band process that has one positive peak for every zero crossing, the proportion of cycles whose peaks exceed $y = a$ is $\nu^+(a)/\nu^+(0)$. This is the probability that any peak chosen at random exceeds a and leads to the result that[1,9]

$$p_p(a) = \frac{a}{\sigma_y^2} \exp\left(-\frac{a^2}{2\sigma_y^2}\right), \qquad a \geq 0, \tag{79}$$

which is the Rayleigh distribution shown in Fig. 2*a*.

6.4 Limit on Spectral Bandwidth

The result (79) depends on the assumption that there is only one positive peak for each zero crossing. It can be shown[10] that Eq. (73) is true for any random process $y(t)$ that can be differentiated at least once, and Eqs. (74) and (78) are true if the process is also Gaussian. If $y(t)$ can be differentiated at least twice, its frequency of maxima will be the same as the frequency of zero crossings of the derived process $\dot{y}(t)$. If the frequency of maxima is μ, then, by analogy with Eq. (78),

$$\mu = \frac{1}{2\pi}\left[\frac{\int_{-\infty}^{\infty} \omega^4 S_{yy}(\omega)\, d\omega}{\int_{-\infty}^{\infty} \omega^2 S_{yy}(\omega)\, d\omega}\right]^{1/2}. \tag{80}$$

Only if the spectral bandwidth is vanishingly small will μ calculated from Eq. (80) be the same as $\nu^+(0)$ calculated from Eq. (78). In other cases there are irregularities in the narrow-band response of Fig. 9 that give rise to additional local peaks.[1,9]

6.5 General Peak Distribution

The probability density function $p_p(a)$ for the distribution of peaks of a general stationary Gaussian process with zero mean, $y(t)$, provided this can be differentiated twice, is[10]

$$\begin{aligned} p_p(a) = {} & \frac{(1-\xi)^{1/2}}{\sqrt{2\pi}\sigma_y} \exp\left\{-\frac{a^2}{2\sigma_y^2(1-\xi^2)}\right\} + \frac{\xi a}{2\sigma_y^2} \\ & \cdot \left[1 + \operatorname{erf}\left\{\frac{\xi a}{\sqrt{2}\sigma_y(1-\xi^2)^{1/2}}\right\}\right. \\ & \left.\cdot \exp\left(-\frac{a^2}{2\sigma_y^2}\right)\right], \end{aligned} \tag{81}$$

where

$$\xi = \frac{\sigma_{\dot{y}}^2}{\sigma_y \sigma_{\ddot{y}}}, \tag{82}$$

which is the ratio of the expected rate of zero crossings to the expected rate of peaks, and erf denotes the error function defined as

$$\operatorname{erf} x = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-w^2)\, dw. \tag{83}$$

In the limiting case of a narrow-band process with vanishingly small bandwidth, there is one peak for each zero crossing so that $\xi = 1$. In that case Eq. (81) reduces to the Rayleigh distribution (79). In contrast, a broadband random process has many peaks for each zero crossing, so that $\xi \rightarrow 0$. In this case $p_p(a)$ is just a normal distribution that is the same as the amplitude distribution.

6.6 Envelope Properties

The *envelope* $A(t)$ of a random process $y(t)$ may be defined in various different ways. For each sample function it consists of a smoothly varying pair of curves such that $|A(t)| \geq |y(t)|$ for all t and for which $|A(t)| = |y(t)|$ at, or nearly at, the peaks of $y(t)$. The various definitions differ in respect of their interpretation of the phrase "or nearly at." One common definition[9] is to say that the envelope is the pair of curves $A(t)$ and $-A(t)$ given by

$$A^2(t) = y^2(t) + \frac{\dot{y}^2(t)}{\nu^+(0)^2}, \tag{84}$$

where $\nu^+(0)$ is the average frequency of zero crossings of the $y(t)$ process.

When $y(t)$ is stationary and Gaussian, it can then be shown that this definition leads to the following envelope probability density function:

$$p(A) = \frac{A}{\sigma_y^2} \exp\left(-\frac{A^2}{2\sigma_y^2}\right), \qquad A \geq 0. \tag{85}$$

This is the same as the probability distribution for the peaks of a narrow-band, stationary, Gaussian process, [Eq. (79)]. However the two distributions differ if the process $y(t)$ is not both narrow band and Gaussian.

6.7 Clumping of Peaks

The response of a narrow-band random process is characterized by a slowly varying envelope, so that peaks

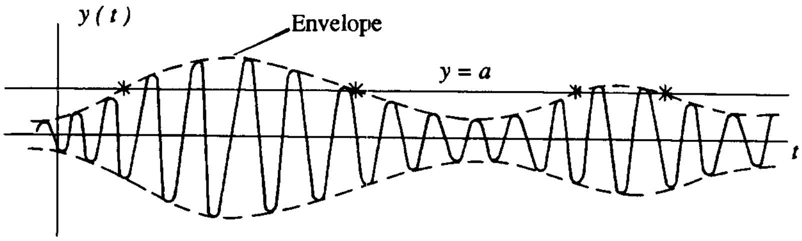


Fig. 10 Each clump of peaks greater than a is begun and ended by an envelope crossing of $y = a$.

occur in clumps (Fig. 10). Each clump of peaks greater than a is begun and ended by its envelope crossing $y = a$. If M denotes the number of up-crossings of $y = a$ of the sample function $y(t)$ for each up-crossing of the same level for its envelope $A(t)$, then

$$E[M] = \frac{\nu_y^+(a)}{\nu_A^+(a)}, \tag{86}$$

which can be shown[10] to be given by

$$E[M] = \frac{\sigma_y}{a}\left[2\pi\left(1 - \frac{\lambda^2}{\sigma_y^2\sigma_{\dot{y}}^2}\right)\right]^{-1/2}, \tag{87}$$

where

$$\lambda = 2\int_0^\infty \omega S(\omega)\,d\omega. \tag{88}$$

For a process whose spectral density is infinitesimally narrow, $\lambda^2 = \sigma_y^2\sigma_{\dot{y}}^2$, and then $E[M]$ becomes infinite as the clumps of peaks are infinitely long. In other cases, $E[M]$ depends on the level a by the factor σ_y/a and on the shape of the spectral density function through the term in square brackets in Eq. (87).

The result (87) does not take account of the fact that a clump cannot have less than one peak. It has been possible to identify and eliminate from the calculation those envelope crossings that are not followed by at least one up-crossing of $y(t)$. This leads to a revised expression for $E[M]$ that is asymptotic to Eq. (87) for narrow-band processes but diverges as the spectral band is widened.

6.8 Nonstationary Processes

The distribution of peaks of a nonstationary random process can also be calculated, and these results then depend on time. If the distribution of $y(t)$ remains Gaussian as it evolves, its peak distribution is close to a Weibull distribution (see Section 2.4) with the parameters X_0 and k [see Eq. (14)] depending on the characteristics of the nonstationary process and on its duration.[10]

The properties of the envelope of a nonstationary process can also be calculated and the level-crossing properties of this envelope may be derived from its probability description.[10]

7 FAILURE DUE TO RANDOM VIBRATION

7.1 Fractional Occupation Time

Failure under vibration may depend on the *fractional occupation time* or *time spent off-range*, which is defined as the proportion Y_a of total time that a random process $y(t)$ spends above the level $y = a$. If $y(t)$ is stationary, Gaussian, and with zero mean, the ensemble average of Y_a is[1]

$$E[Y_a] = \frac{1}{2}\left(1 - \operatorname{erf}\left(\frac{a}{\sqrt{2}\sigma_y}\right)\right). \tag{89}$$

When the level $y = a$ is large, so that $a \geq \sigma_y$, this may be approximated by

$$E[Y_a] \approx \frac{1}{\sqrt{2\pi}}\frac{\sigma_y}{a}\exp\left(-\frac{a^2}{2\sigma_y^2}\right). \tag{90}$$

The variance of Y_a is also available.[10] In addition, both the mean value and variance of Y_a may be calculated when $y(t)$ is nonstationary provided that the probability density functions $p(y(t))$ for the mean value and $p(y(t_1), y(t_2))$ for the variance are available for the total interval of time to be considered.

7.2 First-Passage Time

The *first-passage time* for $y(t)$ is the time at which $y(t)$ first crosses a chosen level, $y = a$, after being started at $t = 0$. If the crossings of level a are a rare event, they may be assumed to be independent of each other and randomly distributed along the time axis. In that case, it can be shown that the probability of a first passage in time T is given by

$$\text{Probability of first passage in time } T = 1 - \exp[-\nu^+(a)T], \qquad (91)$$

were, as in Section 6.2, $\nu^+(a)$ is the average frequency of up-crossings of the level a. The probability density function for first-passage time is obtained by differentiating this expression to obtain

$$p(T) = \nu^+(a)\exp[-\nu^+(a)T], \qquad T > 0, \qquad (92)$$

which gives the following results for the mean and variance of the first-passage time:

$$E[T] = \frac{1}{\nu^+(a)} \quad \text{and} \quad \operatorname{var}[T] = \frac{1}{[\nu^+(a)]^2}. \qquad (93)$$

These results are only correct for crossings randomly distributed along the time axis. Because of clumping, the intervals between clumps will be longer than the average spacing between crossings and the probability of failure in time T will therefore be less than indicated by Eq. (91). The approximation will be particularly poor for a narrow-band process. For this it is more realistic to assume that failure may occur when the envelope $A(t)$ of $y(t)$ first crosses the level a. Assuming that the envelope crossings are independent and have average up-crossing rate $\nu_A^+(a)$, the statistics of the first-passage time are still given by Eq. (93) but with $\nu_A^+(a)$ replacing $\nu^+(a)$.

A general exact solution for the first-passage problem has not yet been found but there are a variety of approximate solutions of considerable complexity. Allowance can be made for $y(t)$ being nonstationary, and upper and lower bounds for the probability of first-passage failure have been calculated.

7.3 Maximum Value

The maximum value of a random process in a fixed interval of time is a random variable. If y_m is the maximum value of a stationary, narrow-band, Gaussian process $y(t)$ in time interval T, then y_m must be the highest peak in interval T. It can be shown that, assuming the peaks are independent, the probability that y_m is less than a is given by

$$P_{y_m}(a) = \exp\left[-\nu^+(0)T\exp\left(-\frac{a^2}{2\sigma_y^2}\right)\right], \qquad (94)$$

where $\nu^+(0)$ is the average frequency of zero crossings of $y(t)$.

The mean and variance of the maximum value y_m can be calculated from $P_{y_m}(a)$ and are given by[10]

$$E[y_m] = \left(C + \frac{0.5772}{C}\right)\sigma_y, \qquad (95)$$

$$\operatorname{var}[y_m] = 1.638\,\frac{\sigma_y^2}{C^2}, \qquad (96)$$

where

$$C = \{2\ln(\nu^+(0)T)\}^{1/2}. \qquad (97)$$

7.4 Fatigue Failure

The calculation of fatigue damage accumulation is based on the notion that individual cycles of stress can be identified and that each stress cycle advances a fatigue crack. When the crack has reached a critical size, failure occurs. One cycle of stress of amplitude S is assumed to generate $1/N(S)$ of the damage needed to cause failure.

For a stationary, narrow-band, random process with average frequency $\nu^+(0)$, the number of cycles in time T will be $\nu^+(0)T$. If the probability density for the distribution of peaks is $p_p(S)$, then the average number of stress cycles in the range S to $S + dS$ will be $\nu^+(0)Tp_p(S)\,dS$.

The damage done by this number of stress cycles is

$$\nu^+(0)Tp_p(S)\,dS\,\frac{1}{N(S)},$$

and so the average damage done by all stress cycles together will be

$$E[D(T)] = \nu^+(0)T\int_0^\infty \frac{1}{N(S)}\,p_p(S)\,dS. \qquad (98)$$

Failure occurs when the accumulated damage $D(T)$ is unity.

The variance of the accumulated fatigue damage, $\sigma_{D(T)}^2$, can also be calculated.[9]

When the stress cycles occur as a result of the response of a lightly damped oscillator to broadband Gaussian excitation, the damping ratio of the oscillator being ζ, it can be shown that

$$\frac{\sigma_{D(T)}}{E[D(T)]} \propto \frac{1}{[\zeta\nu^+(0)T]^{1/2}}. \qquad (99)$$

For light damping, $\zeta \ll 1$, which is necessary for the narrow-band assumption to be correct, the number of stress cycles $\nu^{+}(0)T$ must be large if $\sigma_{D(T)}$ is to be small. In that case, an estimate of the average time to failure, $E[T_F]$, can be made by assuming that this is the value of T in Eq. (98) when $E[D(T)] = 1$.

The practical difficulty with this approach is that of modeling the fatigue crack propagation mechanism sufficiently accurately. The use of more complicated fracture models makes the statistical analysis a good deal more difficult.

8 RESPONSE OF NONLINEAR SYSTEMS

8.1 Introduction

The analysis of how dynamical systems respond to random excitation has been most fully developed for time-invariant, passive, linear systems. Their behavior can be represented by linear differential equations with constant coefficients, and their response is stable, so that a system with no excitation settles eventually to rest. The general excitation–response relationships for such systems are described by the results in Section 5.

In recent years there has been growing interest in the response of parametrically excited linear systems and of nonlinear systems subjected to random excitation. The response of these have not yet been reduced to generally applicable results, and problems of these types are usually solved either by exact methods specific to a particular problem (which, for random vibrations, are rare) or by approximate methods.

In principle, exact solutions for the response of any dynamical system (linear or nonlinear) subjected to broadband, Gaussian excitation (theoretically, white, Gaussian excitation) can be obtained from the theory of *continuous Markov processes*. This method, which is described in Section 8.2, determines the probability structure of a system's response as a function of time. Inherently it is extremely powerful. In practice, the diffusion equation describing how the probability distribution evolves is too difficult to be solved exactly and approximate methods have to be used to obtain a solution. This method forms the basis for one of the main approaches to parametric (time-varying) problems[11] as well as to nonlinear problems. For nonlinear problems there are many different methods, of which two of the more important are *perturbation methods* (Section 8.3) and *statistical linearization* (Section 8.4).

8.2 Markov Methods

For a one-dimensional random process $y(t)$, let $p(y, t|y_0, t_0)\, dy$ be the conditional probability that $y(t)$ has a value between y and $y + dy$ at time $t (t > t_0)$ if it had the value y_0 at time t_0. Then, for a continuous Markov process, the probability density function $p \equiv p(y, t|y_0, t_0)$ satisfies the *Fokker–Planck equation*[10]

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial y}[m(y,t)p] + \frac{1}{2}\frac{\partial^2}{\partial y^2}[\sigma^2(y,t)p], \qquad (100)$$

where $m(y, t)$ is called the *drift coefficient* and $\sigma^2(y, t)$ is the *diffusion coefficient*. These are defined as follows. If Δy is the change in response $y(t)$ that occurs in time increment Δt, then

$$m(y,t) = \lim_{\Delta t \to 0} \frac{E[\Delta y]}{\Delta t} \qquad (101)$$

and

$$\sigma^2(y,t) = \lim_{\Delta t \to 0} \frac{E[\Delta y^2]}{\Delta t}. \qquad (102)$$

For the response of a multidimensional system that has a state vector $y(t)$ with n elements, a corresponding form of Eq. (100) exists with n first-order derivatives $\partial/\partial y_i$ and n^2 second-order derivatives $\partial^2/\partial y_i\, \partial y_j$.

The definition of a *continuous* Markov process[10] requires (i) that the future states of the process $y(t)$ depend only on the present state, not on how that state was arrived at, so that $p(y, t|y_0, t_0)$ is unique, and (ii) that higher order derivate moments like $\lim_{\Delta t \to 0}(E[\Delta y^n]/\Delta t)$ and $\lim_{\Delta t \to 0}(E[\Delta y_i\, \Delta y_j\, \Delta y_k]/\Delta t)$ all vanish when their order is greater than 2.

It can be shown that these conditions always apply when $y(t)$ is the response of a system to excitation that is white and Gaussian.

An alternative version of the Fokker–Planck equation (100) is called the *Kolmogorov equation*, and this expresses the same result but in a slightly different form.

In order to determine the drift and diffusion coefficients in the Fokker–Planck equation, it is necessary to refer to the equations for the dynamical system being studied. For example, consider the one-dimensional (linear) system

$$\dot{y} + cy = x(t), \qquad (103)$$

where $x(t)$ is stationary, white, Gaussian excitation. Since $y(t)$ is continuous but $x(t)$ is not, this may be written in the incremental form

$$\Delta(y) = -cy\,\Delta t + \int_t^{t+\Delta t} x(t)\, dt. \qquad (104)$$

Taking the ensemble average of this equation, subject to the condition that we are considering only samples for which $y(t)$ lies in the range y to $y + \Delta y$, gives

$$\lim_{\Delta t \to 0} \frac{E[\Delta y]}{\Delta t} = -cy \tag{105}$$

because $E[x(t)]$ is zero, and leads to the result that

$$\lim_{\Delta t \to 0} \frac{E[\Delta y^2]}{\Delta t} = 2\pi S_0, \tag{106}$$

where S_0 is the spectral density of $x(t)$.

Exact solutions to the Fokker–Planck equation have been obtained only for a few first-order systems. Even for stationary solutions, occurring as $t \to \infty$ when the time derivative $\partial p/\partial t$ has become zero, there are only solutions for certain specific classes of nonlinear equations.[10] As a result, various methods for obtaining approximate solutions have been developed.

For a lightly damped, one-degree-of-freedom oscillator, the method of *stochastic averaging* allows the two-dimensional Markov process describing the response $y(t)$ to be replaced approximately by a one-dimensional Markov process $a(t)$ describing the envelope of $y(t)$. This method has been used with considerable success for studying first-passage problems.[12] However, its application to systems with more than one degree of freedom has not been successful.

Other approximate methods for solving the *Fokker–Planck* equation include *Galerkin's method,* in which the probability density function is expanded in a set of trial functions that can be calculated approximately by analytical or numerical methods. Also there are *iterative* and *variational* methods that have worked well in particular cases.

8.3 Perturbation Methods

Classical *perturbation methods* have been used successfully to determine the approximate response statistics of weakly nonlinear systems. The basic idea is to expand the solution as a power series in terms of a small scaling parameter. For example, the solution of the system

$$\ddot{y} + \eta\dot{y} + \omega^2 y + \epsilon f(y, \dot{y}) = x(t), \tag{107}$$

where $|\epsilon| \ll 1$, is assumed to have the form

$$y(t) = y_0(t) + \epsilon y_1(t) + \epsilon^2 y_2(t) + \cdots. \tag{108}$$

After substituting Eq. (108) into Eq. (107) and collecting terms, the coefficients of like powers of ϵ are then set to zero. This leads to a hierarchy of linear second-order equations that can be solved sequentially by linear theory. The excitation for the second equation is derived from the response of the first equation, and so on. Using these results, it is then possible to calculate $R_{yy}(\tau)$, for example, since, from Eq. (108),

$$\begin{aligned} R_{yy}(\tau) &= E[y(t)y(t+\tau)] \\ &= R_{y_0y_0}(\tau) + \epsilon\{R_{y_0y_1}(\tau) + R_{y_0y_1}(-\tau)\} + O(\epsilon^2) \end{aligned} \tag{109}$$

and, by taking its Fourier transform,

$$S_{yy}(\omega) = S_{y_0y_0}(\omega) + 2\epsilon\,\mathrm{Re}\{S_{y_0y_1}(\omega)\} + O(\epsilon^2). \tag{110}$$

In practice, results have generally been obtained to first-order accuracy only, because of the complexity of the algebra for higher order calculations. The method can in principle be extended to multi-degree-of-freedom systems but subject to the same problem of algebraic complexity. Provided that $|\epsilon| \ll 1$, results have been found to agree with those from other methods, but a general proof of convergence is not currently available.

8.4 Statistical Linearization

This method involves replacing the governing set of nonlinear differential equations by a set of linear differential equations that are equivalent in some way.[13] In control theory, it is called the *describing function method* and, in random process theory, *equivalent linearization* or *stochastic linearization* as well as *statistical linearization.* The parameters of the equivalent linear system are obtained by minimizing the *equation difference.*

For example, suppose that the linear system that is equivalent to Eq. (107) is

$$\ddot{y} + \eta_e\dot{y} + \omega_e^2 y = x(t). \tag{111}$$

The equation difference is

$$e(y, \dot{y}) = \epsilon f(y, \dot{y}) + (\eta - \eta_e)\dot{y} + (\omega^2 - \omega_e^2), \tag{112}$$

where η_e and ω_e^2 are unknown parameters. They are chosen so as to minimize some measure of $e(y, \dot{y})$. The usual method is to minimize the mean square of $e(y, \dot{y})$ with respect to the parameters η_e and ω_e^2. This is done by setting

$$\frac{\partial E[e^2(y, \dot{y})]}{\partial \eta_e} = 0 \quad \text{and} \quad \frac{\partial E[e^2(y, \dot{y})]}{\partial \omega_e^2} = 0 \tag{113}$$

and leads to two equations for η_e and ω_e^2.

These equations include terms such as $E[y^m \dot{y}^n]$, which are not known unless the joint probability density function $p(y,\dot{y})$ is known, which it is not. This difficulty is overcome by assuming that the probability structure of $y(t)$ and $\dot{y}(t)$ is that of the equivalent system (111) when $x(t)$ is Gaussian. The argument is that, for a lightly damped system, the response $y(t)$ will be approximately Gaussian even if $x(t)$ is non-Gaussian. This enables $E[y^m \dot{y}^n]$ to be obtained in terms of the response statistics of (111), which are expressed in terms of η_e and ω_e^2. On substituting the results into Eq. (113), there are two nonlinear algebraic equations for η_e and ω_e^2. Solving these equations gives η_e and ω_e^2 and the equivalent linear system is determined.

The statistical linearization method can be generalized to multi-degree-of-freedom systems, including those with nonstationary excitation. In the latter case, or when considering the initial response to stationary excitation, the equivalent linear system has time-varying coefficients that are functions of the time-varying response statistics.

There has been considerable research on the statistical linearization method,[13] and it has been applied to many practical response problems, including problems in earthquake engineering with hysteretic damping that occurs due to yielding or slipping. The restoring force is then a function of the previous history of the response, not just its instantaneous state, and the response calculation is more complicated still. Statistical linearization has proved particularly successful in such cases.

8.5 Monte Carlo Simulation

An experimental approach for generating response statistics uses the digital simulation of random processes. Random excitation with the required properties is generated artificially, and the response it causes is found by numerically integrating the equations of motion. Provided that a sufficiently large number of experiments are conducted by generating new realizations of the excitation and integrating its response, so that an ensemble of sample functions is created, response statistics can be obtained by averaging across the ensemble. This permits the statistics of nonstationary responses to be estimated by averaging data from several hundred sample functions. For problems with a stationary response, the necessity for a large number of records can be avoided by averaging over time along a single sample provided that the response statistics required can be assumed to be ergodic.

There are various computational strategies for generating realizations of pseudorandom processes but the most common approach is to use the *inverse discrete Fourier transform*, which is computed by the fast Fourier transform (FFT) algorithm. In order to generate two correlated sample functions $x(t)$ and $y(t)$, with specified spectral densities $S_{xx}(\omega)$, $S_{xy}(\omega) = S_{yx}^*(\omega)$ and $S_{yy}(\omega)$, the procedure is as follows. Over a time interval T, each of $x(t)$ and $y(t)$ is assumed to be replaced by the series of equally spaced discrete values x_r, y_r, $r = 0, \dots, N-1$. According to the *discrete Fourier transform*, there are corresponding (complex) Fourier coefficients X_k, Y_k, $k = 0, \dots, N-1$. These are related to the spectral densities as follows[1]:

$$|X_k|^2 = \frac{2\pi}{T} S_{xx}\left(\omega = \frac{2\pi k}{T}\right),$$
$$|Y_k|^2 = \frac{2\pi}{T} S_{yy}\left(\omega = \frac{2\pi}{T}\right), \qquad (114)$$

and

$$|X_k||Y_k| \exp\{i(\phi_k - \theta_k)\} = \frac{2\pi}{T} S_{xy}\left(\omega = \frac{2\pi k}{T}\right) = \frac{2\pi}{T} S_{yx}^*\left(\omega = \frac{2\pi k}{T}\right), \qquad (115)$$

where ϕ_k, θ_k are the phase angles of the complex coefficients X_k, Y_k.

Therefore the phase differences $(\phi_k - \theta_k)$, $k = 0, \dots, N-1$, are specified but not the absolute values of ϕ_k and θ_k. One or the other, say ϕ_k, is chosen at random using a random-number generator. To generate Gaussian functions, the ϕ_k must be randomly distributed between 0 and 2π. Then, since the sample values x_r, y_r are generated by the inverse discrete Fourier transform relations

$$x_r = \sum_{k=0}^{N-1} X_k \exp\left[i\left(\frac{2\pi kr}{N}\right)\right],$$
$$y_r = \sum_{k=0}^{N-1} X_k \exp\left[i\left(\frac{2\pi kr}{N}\right)\right], r = 0, \dots, N-1, \qquad (116)$$

each term x_r, y_r is the result of summing many statistically independent contributions and so, by the central limit theorem, tends to have Gaussian properties. If a non-Gaussian process has to be simulated, this will generally be done by first generating an artificial Gaussian process and then subjecting the Gaussian process to a nonlinear transformation to obtain the required non-Gaussian properties. Numerical integration of the equations of motion to obtain their response to artificial pseudorandom excitation is generally computed by using one of the Runge–Kutta algorithms.[3,5]

REFERENCES

1. D. E. Newland, *An Introduction to Random Vibrations, Spectral and Wavelet Analysis*, 3rd ed., Longman, Harlow, and Wiley, New York, 1993.
2. R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, 6th ed., Oliver and Boyd, Edinburgh, 1963.
3. D. E. Newland, *Mechanical Vibration Analysis and Computation*, Longman, Harlow, and Wiley, New York, 1989.
4. J. S. Bendat and A. G. Piersol, *Random Data Analysis and Measurement Procedures*, 2nd ed., Wiley, New York, 1986.
5. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes*, Cambridge University Press, Cambridge, 1986.
6. G. M. Jenkins and D. G. Watts, *Spectral Analysis and Its Applications*, Holden-Day, San Francisco, 1968.
7. W. B. Davenport, Jr. and W. L. Root, *An Introduction to the Theory of Random Signals and Noise*, McGraw-Hill, New York, 1958.
8. M. Ohta, K. Hatakeyama, S. Hiromitsu, and S. Yamaguchi, "A Unified Study of the Output Probability Distribution of Arbitrary Linear Vibratory Systems with Arbitrary Random Excitation," *J. Sound Vib.*, Vol. 43, 1975, pp. 693–711.
9. S. H. Crandall and W. D. Mark, *Random Vibration in Mechanical Systems*, Academic Press, New York and London, 1963.
10. N. C. Nigam, *Introduction to Random Vibrations*, MIT Press, Cambridge, MA, 1983.
11. R. A. Ibrahim, *Parametric Random Vibration*, Research Studies Press, Letchworth, U.K., and Wiley, New York, 1985.
12. J. B. Roberts and P. D. Spanos, "Stochastic Averaging: An Approximate Method for Solving Random Vibration Problems," *Int. J. Non-Linear Mech.*, Vol. 21, 1986, pp. 111–134.
13. J. B. Roberts and P. D. Spanos, *Random Vibration and Statistical Linearization*, Wiley, Chichester, 1990.

67

SHOCK ANALYSIS AND DESIGN

THOMAS L. GEERS

1 INTRODUCTION

A shock excitation is *transient*, that is, it has a beginning and (usually) an end. Mechanical shock is conveyed to a mechanical system through forces and motions applied at one or more points. Most often, the system is in a quiescent state when the shock is applied; hence, the response of the system to the shock is also transient. Generally, the aspect of greatest interest in a shock problem is the possibility of system damage or failure.

Damage or failure in a mechanical system is caused by internal stresses and/or relative displacements that exceed the system's capacity to tolerate them. Thus, a shock *analysis* focuses on determining the peak responses that might produce damage and/or failure, and a shock *design* concentrates on controlling those peak responses. The central importance of response peaks that are often highly localized in both time and space constitutes a challenging aspect of shock problems. An equally challenging aspect is relating the response peaks to actual damage or failure.

The chapter is organized as follows. The next four sections focus on *shock analysis*, with Sections 2 and 3 discussing equations of motion for mathematical models and Sections 4 and 5 addressing the analytical and numerical solution of those equations. The following four sections focus on *shock design*, with Sections 6 and 7 describing design tools and Sections 8 and 9 discussing design practices. The chapter continues with two sections on the special topic of transient structure–medium interaction and concludes with a few remarks on the state of the technology.

2 FORMULATION OF SHOCK PROBLEMS

In the vast majority of cases, shock problems are mathematically formulated in terms of time-dependent differential equations of motion. If the system model is *discrete* (characterized by lumped elements such as masses, springs, and dampers), these equations are ordinary differential equations in time. They are supplemented with a set of initial conditions that specify the state of the system just before the shock is applied. If the system model is *continuous* (characterized by spatially varying material properties such as mass density and Young's modulus), the equations of motion are partial differential equations in time and space. They are supplemented with initial conditions that specify the state of the system over all space just before the shock is applied and with boundary conditions that specify the state of the system over all time at given boundaries of the system.

A surprising number of shock problems may be satisfactorily formulated in terms of the single-degree-of-freedom, linear oscillator of Fig. 1. The equation of motion for this model when both force and displacement excitations are present is[1]

$$m\ddot{u}(t) + b\dot{u}(t) + ku(t) = f(t) + b\dot{w}(t) + kw(t), \qquad (1)$$

where m, b, and k are the oscillator mass, damping constant, and spring constant, respectively, $u(t)$ is the displacement of the oscillator mass from its static equilibrium position, $f(t)$ is the transient external force acting on the oscillator mass, and $w(t)$ is the transient external displacement acting at the base of the oscillator; an overdot denotes an ordinary differentiation in t. This ordinary differential equation must be supplemented with the initial conditions $u(0) = u_0$ and $\dot{u}(0) = \dot{u}_0$, where, without loss of generality, $t = 0$ is taken as the time just before

---

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker

Copyright © 1997 John Wiley & Sons, Inc.

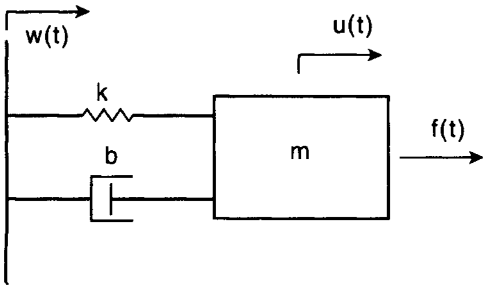


Fig. 1 Linear oscillator subjected to external force and base-displacement excitations.

the shock is applied; usually quiescent initial conditions apply, that is, u_0 and $\dot{u}_0$ are both zero. Many mechanical systems require multi-degree-of-freedom representations, which may be produced by direct modeling or by finite-difference or finite-element models of continuous systems. In such cases, Eq. (1) takes the form of a matrix equation, with m, b, and k becoming matrices and $u(t)$, $f(t)$, and $w(t)$ becoming computational column vectors.

At times, a continuous model is the most appropriate representation of a shock-excited system. A simple example of such a model is the undamped elastic bar of Fig. 2, the equation of motion for which is[2]

$$\rho(x)A(x)\ddot{u}(x,t) - [E(x)A(x)u'(x,t)]' = f(x,t), \quad (2)$$

where $\rho(x)$ and $E(x)$ are, respectively, the mass density and Young's modulus for the bar material, $A(x)$ is its cross-sectional area, $u(x,t)$ is the displacement of an infinitesimally thin disc of bar material from its static equilibrium position, and $f(x,t)$ is the transient external force distribution acting along the bar; a dot or prime denotes a partial differentiation in t or x, respectively. This partial differential equation must be supplemented with the initial conditions $u(x,0) = u_0(x)$ and $\dot{u}(x,0) = \dot{u}_0(x)$. It must also be supplemented with boundary conditions, which, for the configuration shown in Fig. 2, are $u(0,t) = w(t)$ and $u'(l,t) = 0$. There are even times when a mixed discrete/continuous model is the most appropriate representation: an example would be the elastic bar with a lumped mass on the end.

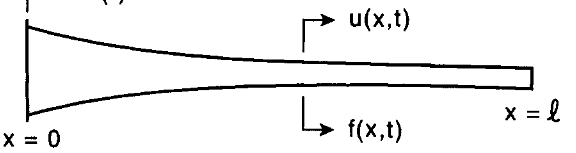


Fig. 2 Elastic bar subjected to external distributed-force and base-displacement excitations.

Equations (1) and (2) constitute equations of motion for *linear* models of mechanical systems. In many cases, nonlinear system behavior determines survivability. *Constitutive nonlinearity* typically appears in Eq. (1) when $k = k[r(t), \dot{r}(t)]$, where $r(t)$ is the relative displacement $u(t) - w(t)$; clearly, the spring "constant" is no longer constant. It typically appears in Eq. (2) when $E = E[x, \epsilon(x,t), \dot{\epsilon}(x,t)]$, where $\epsilon(x,t) = u'(x,t)$ is longitudinal strain. Constitutive nonlinearity is much more complicated in two and three dimensions. *Geometric nonlinearity* appears when motions exceed certain tolerances, such as the geometric limits of linear deformation and the "rattlespace" existing in the system. A rattlespace violation would occur in the system of Fig. 1 if the mass contacted another body. The appearance of nonlinear behavior in a previously linear system fundamentally alters system response and greatly complicates analysis and design (see Chapter 65). Fortunately, dynamic finite-element computer codes that treat both constitutive and geometric nonlinearity are widely available.

3 SHOCK EXCITATIONS

As indicated above, a shock excitation $e(t)$ can be quite general, constrained only by the fact that it must be transient, which can be stated as $e(t) = 0$ for $t < 0$. A variety of *idealized excitations* are useful in both theory and practice, such as rectangular, triangular, and exponential pulses, which are pictured below in Fig. 5, as well as the half-sine pulse, given by $A \sin(\pi t/T)$ for $0 \le t \le T$ and zero otherwise.

Two idealized excitations are especially useful. The first of these is the Heaviside step function[3] $H(t - \tau)$, where $\tau \ge 0$. The step function is equal to zero for $t < \tau$ and to unity for $t > \tau$; its value at $t = \tau$ is not precisely defined but may be conveniently taken as $\frac{1}{2}$. Note that any function, when multiplied by $H(t)$, becomes a transient function; for example, a pure sinusoid, when multiplied by $H(t - \tau)$, becomes a step sinusoid starting at $t = \tau$.

The second is the Dirac delta function[3] $\delta(t - \tau)$, which is the conceptual derivative of $H(t-\tau)$. The delta function is equal to zero for both $t < \tau$ and $t > \tau$ and is equal to infinity at $t = \tau$; furthermore, the area beneath it is unity. With these unconventional characteristics, the delta function is best viewed as the limit of a smoother function characterized by a parameter that can make the function simultaneously higher and narrower while maintaining unit area. An example is the rectangular pulse of width Δt and height $1/\Delta t$ with $\Delta t \to 0$; another is the isosceles triangular pulse with base Δt and height $2/\Delta t$ with $\Delta t \to 0$. Note that the physical dimension of $\delta(t - \tau)$ is reciprocal time. The most valuable characteristic of the delta function is its *filtering property*:

$$\int_0^\infty \delta(t-\tau)v(\tau)\,d\tau = v(t), \qquad (3)$$

where $v(t)$ is any smooth function.

4 ANALYTICAL SOLUTION OF SHOCK PROBLEMS

When the system and excitation are sufficiently simple, shock response solutions may be obtained analytically. A useful mathematical tool in this regard is the *Laplace transform*,[3] defined by

$$V(s) = L[v(t)] \equiv \int_0^\infty v(t)e^{-st}\,dt, \qquad (4)$$

in which the transform variable s is complex, that is, $s = s_R + is_I$, where s_R and s_I are real with physical dimension of reciprocal time. Laplace transformation is useful because it transforms differential expressions into algebraic expressions; for example, it is easily shown through integration by parts that $L[\dot{v}(t)] = sV(s) - v(0)$ and $L[\ddot{v}(t)] = s^2V(s) - sv(0) - \dot{v}(0)$. Thus Eq. (1) transforms from an ordinary differential equation to an algebraic equation and Eq. (2) from a partial differential equation to an ordinary differential equation; note that initial conditions are automatically incorporated. Solution is thereby simplified, provided that inverse Laplace transformation may be subsequently performed with reasonable effort.

As an example of solution by Laplace transformation, consider Eq. (1) for $w(t) = 0$ and quiescent initial conditions (see Fig. 1); transformation yields

$$U(s) = \frac{F(s)}{s^2m + sb + k}. \qquad (5)$$

Thus, one may transform a known $f(t)$ to get $F(s)$, insert the result into Eq. (5), and inverse transform to find $u(t)$. Inverse transformation may be done analytically or through the use of widely available tables. For instance, consider the excitation $f(t) = \delta(t)$, for which, from Eqs. (4) and (3), $F(s) = 1$; in this case, $U(s)$ in Eq. (5) becomes the *transfer function* for the linear oscillator, and its inverse transform, the *impulse response*, is given for $b < 2\sqrt{mk}$ by

$$h(t) = (m\Omega_n)^{-1}e^{-\zeta\omega_n t}\sin\Omega_n t, \qquad (6)$$

where the *undamped natural frequency* $\omega_n = \sqrt{k/m}$, the *damping ratio* $\zeta = b/2m\omega_n$, and the *damped natural frequency* $\Omega_n = \sqrt{1-\zeta^2}\omega_n$. For a more extensive discussion of this topic, see Chapter 63.

Another useful property of the Laplace transform is that *multiplication* in the s-domain corresponds to *convolution* in the t-domain (Borel's theorem). Thus, for any transient force $f(t)$, one can write from Eq. (5) the displacement solution in the form of the convolution integral

$$u(t) = \int_0^t h(t-\tau)f(\tau)\,d\tau. \qquad (7)$$

This formula portrays the response $u(t)$ as a superposition of impulse responses, each delayed in time and weighted by the value of the force at the appropriate time. The convolution relation (7) is often expressed formally as $u(t) = h(t) * f(t)$.

As a second example, consider Eq. (2) with $\rho(x) = \rho_0$, $A(x) = A_0$, $E(x) = E_0$, and $f(x,t) = 0$, with initial conditions $u(x,0) = 0$, $\dot{u}(x,0) = 0$, and with boundary conditions $u(0,t) = w(t)$, $u'(l,t) = 0$ (see Fig. 2). Of course, $w(t) = 0$ for $t < 0$. Laplace transformation then yields, with $c = \sqrt{E_0/\rho_0}$ as the speed of sound in the bar,

$$c^2U''(x,s) - s^2U(x,s) = 0, \qquad (8)$$

the solution to which is $U(x,s) = C_+(s)e^{sx/c} + C_-(s)e^{-sx/c}$. Here, $C_+(s)$ and $C_-(s)$ follow from the boundary conditions, which yield $C_+(s) + C_-(s) = W(s)$ and $C_+(s) = C_-(s)e^{-2sl/c}$. Yet another useful property of the Laplace transform is the *shifting property* $L[v(t-\tau)] = V(s)e^{-s\tau}$. Thus, the solution for $u(x,t)$ in this example may be obtained by inverse transformation as

$$u(x,t) = w\left(t - \frac{x}{c}\right) + w\left(t + \frac{x}{c} - \frac{2l}{c}\right) - u\left(x, t - \frac{2l}{c}\right). \qquad (9)$$

This is a *traveling-wave* solution, with the first term on the right defining a wave propagating to the right and the second term a delayed wave propagating to the left; the last term subtracts from the two travelling waves the displacement field that existed at time $2l/c$ earlier.

It is interesting to compare Eq. (9) with the *standing-wave* solution for this problem.[4] To obtain this solution, it is convenient to work with the relative displacement $r(x,t) = u(x,t) - w(t)$, for which Eq. (8) yields

$$c^2R''(x,s) - s^2R(x,s) = s^2W(s). \qquad (10)$$

One begins by setting $W(s) = 0$ in this equation, introducing the change of variable $s = i\omega$, and solving the resulting equation to obtain $R(x, i\omega) = S(i\omega)\sin(\omega x/c) + C(i\omega)\cos(\omega x/c)$. The terms $S(i\omega)$ and $C(i\omega)$ follow from the *homogeneous* boundary conditions $r(0,t) = 0$ and $r'(l,t) = 0$, which yield $C(i\omega) = 0$ and $\cos(\omega l/c) = 0$. The latter generates an infinite number of *natural frequencies* $\omega_n = n\pi c/2l$ for $n = 1, 3, 5, \ldots$. Thus, returning to the transform variable s, one can express the solution for $R(x,s)$ in terms of the *mode shapes* $\sin(\omega_n x/c) = \sin(n\pi x/2l)$ as the *modal sum*

$$R(x,s) = \sum_{n=1,3,5,\ldots}^{\infty} R_n(s)\sin\frac{n\pi x}{2l}. \qquad (11)$$

The unknown coefficients $R_n(s)$ are determined through exploitation of the *orthogonality property* of the mode shapes:

$$\int_0^l \sin\frac{m\pi x}{2l}\sin\frac{n\pi x}{2l}\,dx = \begin{cases} 0, & m \neq n, \\ \frac{1}{2}l, & m = n \end{cases}. \qquad (12)$$

One does this by introducing Eq. (11) into Eq. (10), multiplying both sides of the resulting equation by $\sin(m\pi x/2l)$, integrating over the length of the bar, and employing Eq. (12) to obtain

$$R_n(s) = -\frac{4}{n\pi}\,\frac{s^2 W(s)}{s^2 + \omega_n^2}. \qquad (13)$$

Thus, introduction of the Laplace transform of $w(t)$ into Eq. (13) and inverse transformation of the result yields $r_n(t)$. The solution for $u(x,t) = w(t) + r(x,t)$ then follows from Eq. (11) as

$$u(x,t) = w(t) + \sum_{n=1,3,5,\ldots}^{\infty} r_n(t)\sin\frac{n\pi x}{2l}. \qquad (14)$$

The solution may be neatly expressed in convolution form by writing $s^2/(s^2 + \omega_n^2) = 1 - \omega_n^2/(s^2 + \omega_n^2)$, so that inverse transformation of Eq. (13) gives $r_n(t) = -(4/n\pi)[w(t) - \omega_n \sin\omega_n t * w(t)]$; when introduced into Eq. (14), this yields

$$u(x,t) = 2\frac{c}{l}\sum_{n=1,3,5,\ldots}^{\infty}\left[\sin\left(\frac{n\pi c}{2l}\right) t * w(t)\right]\sin\frac{n\pi x}{2l}. \qquad (15)$$

For further discussion of solution by modal superposition see Section 4.3 of Chapter 62, Section 2.2 of Chapter 63, and Section 2.9 of Chapter 64.

Figure 3 contains snapshots of the velocity field $\dot{u}(x,t)$ produced by Eqs. (9) and (14) for a ramp displacement input $w(t) = v_0 t$, for which Eq. (13) yields $r_n(t) = -(4v_0/n\pi\omega_n)\sin\omega_n t$; the series in Eq. (14) is truncated after 10 terms ($n_{\max} = 19$). It is interesting that the standing-wave, or modal-summation, solution cannot reproduce the jump in the exact traveling-wave solution, generating instead spurious oscillations near the wave-

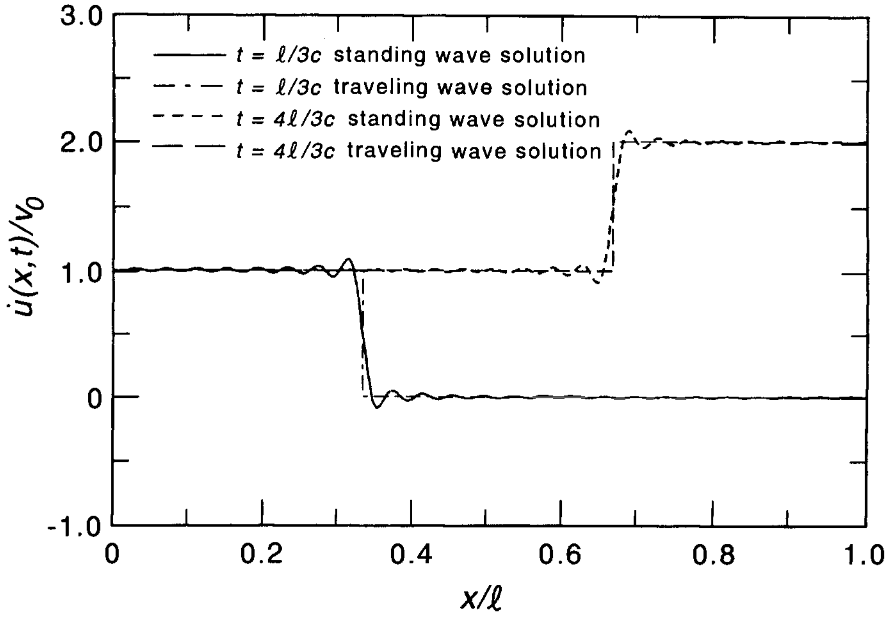


Fig. 3 Velocity-field snapshots in a uniform bar for a ramp-displacement base excitation.

front. This artifact, known as Gibbs's phenomenon,[5] cannot be eliminated by increasing the value of $n_{\max}$; doing so pulls the oscillations in closer to the jump, but does not reduce their magnitude. From Eq. (9), strain snapshots $\epsilon(x,t) = u'(x,t)$ and stress snapshots $\sigma(x,t) = E_0 u'(x,t)$ would exhibit the same characteristics.

5 NUMERICAL SOLUTION OF SHOCK PROBLEMS

Numerical methods offer a means for solving shock problems when systems and/or excitations are complicated. The approach is that so often followed to obtain analytical solutions, namely, reducing differential equations to algebraic equations.

For discrete-system models, the equations of motion are typically ordinary differential equations in time, as discussed in Section 2. For continuum models, the finite-element method[6] is widely used to reduce the partial differential equations of motion to ordinary differential equations. For example, the displacement field in the bar of Fig. 2 is approximated as $u(x,t) \approx N(x)u(t)$, where $N(x)$ is a row vector of spatially local shape functions and $u(t)$ is a column vector of nodal responses. Then a variational principle or weighted-residual method is applied to yield Eq. (1) with m as the mass matrix, b as the damping matrix, k as the stiffness matrix, $f(t)$ as the external force vector, and $w(t)$ as the external displacement vector. The systematic process of reducing spatially continuous models of dynamic systems to spatially discrete models is called *semidiscretization*. Because semidiscretization is akin to standing-wave solution, finite-element solutions for traveling waves with abrupt wavefronts exhibit Gibb's phenomenon.

For discrete and semidiscretized models, the ordinary differential equations in time (which may be thousands in number) are usually integrated by means of finite-difference schemes.[7] With the time continuum sampled evenly over one or more intervals at points $t_j = jh$, where h is a small time increment (not an impulse response!), response histories are obtained within each interval as $u(jh) = u_j$.

One of the most popular time-integration schemes is the *central-difference* scheme, in which displacement and velocity at discrete points in time are approximated as $u_j \approx u_{j-1} + h\dot{u}_{j-1/2}$, $\dot{u}_j \approx \frac{1}{2}(\dot{u}_{j+1/2} + \dot{u}_{j-1/2})$, and $\dot{u}_{j+1/2} \approx \dot{u}_{j-1/2} + h\ddot{u}_j$. These may be combined with Eq. (1) expressed at time t_j to obtain

$$u_{j+1} \approx D^{-1}[h^2 g_j + (2m - h^2 k)u_j - (m - \tfrac{1}{2}hb)u_{j-1}], \quad (16)$$

where $D = (m + \frac{1}{2}hb)$, $g_j = f_j + b\dot{w}_j + kw_j$, and D^{-1} denotes matrix inversion of D, which is implemented in large computations through matrix factorization. Discrete response histories are obtained by advancing j in unit steps, with starting points determined from the initial conditions as $u_0 = u_0$ and $u_{-1} = u_0 - h\dot{u}_0 + \frac{1}{2}h^2\ddot{u}_0$, where $\ddot{u}_0 = m^{-1}(g_0 - b\dot{u}_0 - ku_0)$. This scheme works quite well as long as the selected time increment is no larger than the critical time increment $h_c = 2/\omega_N$, where ω_N is the largest natural frequency in the discrete system; when $h > h_c$, the scheme is numerically unstable. Such conditional stability is the hallmark of an *explicit* numerical integration scheme. A group of equally popular explicit schemes is the family of Runge–Kutta algorithms.

In contrast to explicit schemes, some (but not all) *implicit* schemes are unconditionally stable with respect to time increment. Perhaps the most popular of these is the *trapezoidal rule*, which is one member of the family of Newmark algorithms. The trapezoidal rule approximations for displacement and velocity are $u_{j+1} \approx u_j + \frac{1}{2}h(\dot{u}_j + \dot{u}_{j+1})$ and $\dot{u}_{j+1} \approx \dot{u}_j + \frac{1}{2}h(\ddot{u}_j + \ddot{u}_{j+1})$, which may be combined with Eq. (1) expressed at t_{j+1} and t_j to obtain

$$\begin{aligned} u_{j+1} \approx E^{-1}[&\tfrac{1}{4}h^2(g_{j+1} + 2g_j + g_{j-1}) + (m + \tfrac{1}{2}hb - \tfrac{1}{4}h^2k) \\ &\cdot (u_j + u_{j-1}) + 2m(u_j - u_{j-1})] - u_j, \end{aligned} \quad (17)$$

where $E = m + \frac{1}{2}hb + \frac{1}{4}h^2k$. Discrete response histories are obtained by advancing j in unit steps, with starting points determined from the initial conditions, as discussed above.

Equations (16) and (17) indicate that the implicit trapezoidal rule is the better choice for integrating linear equations of motion with few changes in time increment, in that the costs of securing unconditional stability are a modest increase in algorithmic complexity and infrequent inversion (factorization) of a more complicated matrix. They indicate that the explicit central-difference scheme is the better choice for integrating linear equations with frequent time increment changes when $b = 0$ or when m and b are diagonal and for integrating equations with stiffness nonlinearity; in these situations, matrix inversion (factorization) is minimized by the central-difference method.

Some shock problems consist primarily of high-frequency response over a short period of time and tend to exhibit traveling-wave behavior. Others consist primarily of low-frequency response over a long period of time and tend to exhibit standing-wave behavior. Still other problems exhibit both types of behavior, for example, when a short, rapidly applied load produces long-lived free vibration. An explicit time-integration scheme is indicated for high-frequency problems because the time increment required for accurately tracking the response may be smaller than that required for numerical stability. An implicit scheme is indicated for low-frequency problems because the time increment required for accuracy may be far larger than that required for stability of explicit integration; problems of this type are called *stiff problems*.

Often, the excitation employed to perform a shock response calculation is a *standard pulse form*[8] that represents in a general sense the anticipated shock environment. An example is the half-sine pulse, which is characterized by a maximum value and a pulse duration. More complicated forms involve several parameters that control rise time, maximum value, attenuation rate, and oscillatory behavior. Parameter values are usually selected on the basis of experience with existing experimental data.

6 SHOCK SPECTRUM

Designers are often less interested in obtaining the solution to a particular shock problem than in characterizing a shock *environment* in such a way that system response is easily estimated. The most widely used design tool for accomplishing the latter is the *shock spectrum*,[9] which is a graph that characterizes a shock *excitation* in terms of the peak response to that excitation of an undamped linear oscillator. Although applicable to force excitations, the shock spectrum is usually employed to characterize motion excitations in terms of the oscillator's relative displacement $r(t) = u(t) - w(t)$.

To construct a shock spectrum for a motion excitation $w(t)$, one determines the maximum absolute value of $r(t)$ as a function of the natural frequency $\omega_n = \sqrt{k/m}$ from the following form of Eq. (1) with $b = f(t) = 0$:

$$\ddot{r}(t) + \omega_n^2 r(t) = -\ddot{w}(t). \tag{18}$$

Replacing the right side of this equation with $\delta(t)$, one finds an impulse response given by Eq. (6) with $m = 1$ and $\zeta = 0$; hence, from Eq. (7),

$$\begin{aligned} r(t) &= -\omega_n^{-1} \int_0^t \sin \omega_n(t-\tau)\, \ddot{w}(\tau)\, d\tau \\ &= \omega_n^{-1}[C(\omega_n, t) \cos \omega_n t - S(\omega_n, t) \sin \omega_n t], \end{aligned} \tag{19}$$

where

$$\begin{aligned} C(\omega_n, t) &= \int_0^t \ddot{w}(\tau) \sin \omega_n \tau\, d\tau, \\ S(\omega_n, t) &= \int_0^t \ddot{w}(\tau) \cos \omega_n \tau\, d\tau. \end{aligned} \tag{20}$$

The solution is conveniently plotted as a *pseudovelocity* shock spectrum $V_{Ps}(\omega_n) = \omega_n |r(t)|_{\max}$, that is,

$$V_{Ps}(\omega_n) = [C^2(\omega_n, t) + S^2(\omega_n, t)]^{1/2}_{\max}. \tag{21}$$

With $V_{Ps}(\omega_n)$ thus determined for a given motion excitation, one can immediately calculate, for a mass–spring oscillator subjected to that excitation, the maximum relative displacement across the spring as $|r(t)|_{\max} = \omega_n^{-1} V_{Ps}(\omega_n)$ and, from Eq. (18), the maximum acceleration of the oscillator's mass as $|\ddot{u}(t)|_{\max} = \omega_n V_{Ps}(\omega_n)$. The inclusion of oscillator damping spoils this symmetry, but because $\zeta \ll 1$ in most mechanical systems, the effects of damping are usually neglected.

The pseudovelocity shock spectrum for a rectangular acceleration pulse described by $\ddot{w}(t) = A[H(t) - H(t-T)]$ is readily determined as

$$V_{Ps}^{\text{rect}}(\omega_n) = AT \begin{cases} (\frac{1}{2}\omega_n T)^{-1} \sin \frac{1}{2}\omega_n T, & 0 \le \omega_n T \le \pi, \\ (\frac{1}{2}\omega_n T)^{-1}, & \omega_n T \ge \pi. \end{cases} \tag{22}$$

This function is plotted in Fig. 4. Also plotted is the magnitude of the pulse's Fourier transform,[10] given by

$$\mathcal{F}_{\text{Mag}}^{\text{rect}}(\omega) = AT(\tfrac{1}{2}\omega T)^{-1} |\sin \tfrac{1}{2}\omega T|, \tag{23}$$

where the circular frequency ω is the Fourier transform variable. Note that, even though the Fourier transform vanishes at $\omega T = 2\pi, 4\pi, \ldots$, oscillators with $\omega_n T = 2\pi, 4\pi, \ldots$ respond fully to the excitation. This demonstrates that a shock spectrum is different from the magnitude of a Fourier transform.

The format of Fig. 4 clearly lends itself to quick assessment of peak relative-displacement and absolute-acceleration values. The shock spectrum for the rectangular acceleration pulse approaches a line of constant velocity (AT) at low natural frequencies and a line of constant acceleration ($2A$) at high natural frequencies. In contrast, the shock spectrum for an excitation characterized by a finite maximum displacement approaches a constant-displacement line at low natural frequencies; if the excitation is also free of acceleration jumps and characterized by a peak acceleration, its shock spectrum approaches a constant-acceleration line at high natural frequencies. These more realistic characteristics form the basis of *shock design spectra*, an example of which is included in Fig. 4.

In many (if not most) situations, the system of interest is actually a subsystem attached to a larger system and $w(t)$ is the motion at the interface between the two. If the subsystem is small relative to the larger system, then a shock spectrum based on the motion of the inter-

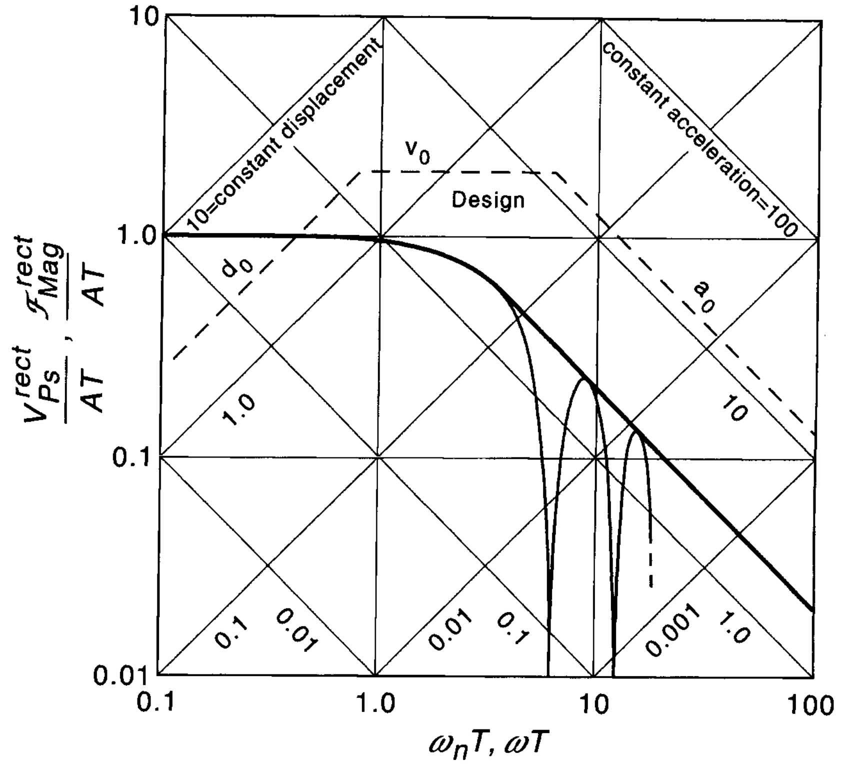


Fig. 4 Pseudovelocity shock spectrum (thick line) and Fourier magnitude spectrum (thin line) for a rectangular acceleration pulse; representative shock design spectrum (dashed line).

face in the absence of the subsystem is appropriate. If this is not the case, however, the subsystem can generate considerable feedback, thereby modifying the interface motion. This produces what is known as the *shock spectrum dip*,[11] which is produced by the subsystem's high resistance to base motion at excitation frequencies near ω_n. The result is that, in the vicinity of ω_n, $V_{Ps}(\omega_n)$ decreases with increasing oscillator mass in a complicated way that depends on the mechanical characteristics of the larger system. Nevertheless, it is often possible to generate shock design spectra that are representative of the general environment, either from experimental data or from numerical simulations.[12]

A limitation of the shock spectrum is the difficulty of extending it to multi-degree-of-freedom linear systems. A modal analysis of a base-excited system permits one to represent it as an array of parallel, uncoupled modal oscillators resting on the base[12]; the peak response of each modal oscillator then follows from the shock spectrum for the motion excitation. However, the maximum response at any physical point in the system is a superposition of modal response contributions at just the right time, not a sum of modal shock-spectrum contributions. There exists an upper bound for maximum response at a physical point as a sum of modal shock spectrum contributions, but it is generally far too conservative. A popular approximation is to determine from the shock spectrum the maximum contribution of each mode to the point response of interest and then to estimate the maximum response there as the vector sum (i.e., the square root of the sum of the squares) of the modal contributions.

For example, suppose one wished to estimate, by vector summation of modal strain contributions, the maximum strain at the root of the bar in Fig. 2 when it is excited by the base motion $w(t) = v_0 t$. Introduction of this motion's Laplace transform into Eq. (13), multiplication of both sides of the resulting equation by $(s^2+\omega_n^2)$, and inverse transformation yield

$$\ddot{r}_n(t) + \omega_n^2 r_n(t) = -\frac{4v_0}{n\pi}\,\delta(t). \tag{24}$$

But this is identical to Eq. (18) with $r(t) = r_n(t)$ and $\ddot{w}(t) = (4v_0/n\pi)\delta(t)$. Thus, Eqs. (20) and (21) yield $C(\omega_n, t) = 0$, $S(\omega_n, t) = 4v_0/n\pi$, $V_{Ps}(\omega_n) = 4v_0/n\pi$, and $|r_n(t)|_{\max} = \omega_n^{-1} V_{Ps}(\omega_n) = 8v_0 l/(n\pi)^2 c$.

Now from Eq. (14), the modal superposition expression for the strain field is

$$\epsilon(x,t) = u'(x,t) = \sum_{n=1,3,5,\dots}^{\infty} \frac{n\pi}{2l}\, r_n(t) \cos\frac{n\pi x}{2l}. \tag{25}$$

Thus, the maximum modal strain at the base is $|\epsilon_n(0,t)|_{\max} = (n\pi/2l)|r_n(t)|_{\max} = 4v_0/n\pi c$, and the vector sum of the modal strain contributions is

$$\bar{\epsilon}(0) = \frac{4v_0}{\pi c}\left(\sum_{n=1,3,5,\ldots}^{\infty} n^{-2}\right)^{1/2} = \sqrt{2}\,\frac{v_0}{c}. \qquad (26)$$

This is 41% higher than the exact value of v_0/c, which may be obtained from Eq. (9).

Another limitation is the difficulty of extending the shock spectrum concept to nonlinear systems. For a single-degree-of-freedom model, the difficulty is not too great, once a means for characterizing the nonlinearity has been established. The system parameters required to describe different types and degrees of nonlinearity tend to proliferate, however, and the concept loses its simplicity of application. Further extension into nonlinear multi-degree-of-freedom systems appears impractical.

7 ISORESPONSE CURVE

Another shock design tool that should be mentioned in the isoresponse curve[13]; which is not restricted either to single-degree-of-freedom systems or to linear systems. To construct an isoresponse curve for *force excitation of a particular system*, one selects a system response quantity of interest and a specific type of loading. The peak force $F = \max|f(t)|$ and the force impulse $I = |\int_0^\infty f(t)\,dt|$ of the loading are allowed to vary. Computations are carried out, or experiments are performed, that enable one to determine several combinations of F and I that produce a certain peak level of the selected system response.

For example, consider the displacement response of an undamped linear oscillator (Fig. 1) subjected to rectangular-pulse forcing. The equation of motion is Eq. (1) with $b = w(t) = 0$. For an infinitely short rectangle with impulse $I_\delta, f(t) = I_\delta \delta(t)$. This produces $u(t) = I_\delta h(t)$, where $h(t)$ is given by Eq. (6) with $\zeta = 0$; peak displacement response is therefore $|u(t)|^\delta_{\max} = I_\delta/m\omega_n$. For an infinitely long rectangle with force value $F_H, f(t) = F_H H(t)$. This produces $u(t) = (F_H/m\omega_n^2)(1 - \cos\,\omega_n t)$; peak displacement response is therefore $|u(t)|^H_{\max} = 2F_H/m\omega_n^2$. Thus, to obtain a certain peak displacement of interest $|u|_{\max}$, the required delta-loading impulse is $I_\delta = m\omega_n|u|_{\max}$ and the required step-loading force is $F_H = \frac{1}{2}m\omega_n^2|u|_{\max}$.

Now the peak displacement response of the undamped linear oscillator to a rectangular loading of finite magnitude F and finite duration T is readily determined. Thus, for a particular value of $I = FT > I_\delta$, it is obviously possible to find the corresponding value of $F > F_H$ that produces $|u(t)|_{\max} = |u|_{\max}$. This exercise can be repeated for several different values of I to obtain an ensemble of isoresponse I, F pairs, of which $I = I_\delta$, $F = \infty$ and $I = \infty$, $F = F_H$ are two members. Finally, all the I-values and F-values are normalized to I_δ and F_H, respectively, and the locus of the normalized isoresponse pairs is plotted to obtain the isoresponse curve marked "rectangular load" in Fig. 5. As indicated, the other curves in this figure are isoresponse curves for triangular and exponential loadings. Because I_δ and F_H *are independent of the loading profile*, Fig. 5 shows that the rectangular (exponential) profile is the most brutal (gentle) of the three.

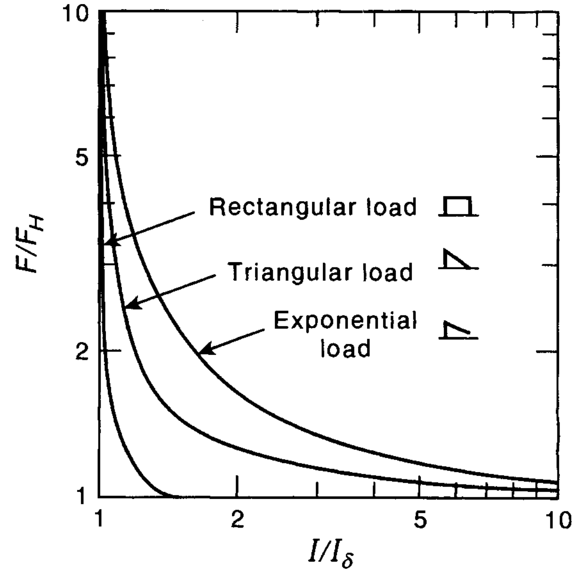


Fig. 5 Isoresponse curves for the peak displacement of an undamped linear oscillator excited by external force pulses. (From G. R. Abrahamson and H. E. Lindberg, *Nuclear Engineering and Design*, Vol. 37, 1976, pp. 35–46. Copyright ©1976 Elsevier Science S. A. Used with permission.)

Isoresponse curves have been generated for a variety of linear and nonlinear systems of widely varying complexity. Analytical, computational, and experimental techniques have all been used.

8 SHOCK SURVIVAL

The design of a mechanical system to survive a shock environment generally employs one or more of the following three strategies: *hardening the system*, *mitigating the shock*, and *increasing the rattlespace*. Hardening the system involves obvious measures such as reducing stress by enlarging cross-sectional areas, smoothing corners and utilizing tougher materials; brittle materials are adverse to shock hardness. One measure that is not so obvious consists of increasing material volume to accommodate greater strain energy. For example, a bolt in uniform tension can hold a maximum elastic strain energy given by $\frac{1}{2}E_0A_0l\epsilon_y^2$, where E_0 and A_0 are the uniform bar properties discussed in Section 4, l is the length of the bolt, and ϵ_y is the yield strain of the bolt material. Thus, a designer might alter the hold-down configuration of a system in order to incorporate longer bolts, thereby increasing their shock hardness.

The most common type of shock mitigation is cushioning. Suppose, for example, that a designer wishes to reduce the peak acceleration of the mass in Fig. 1 to a level below that characterizing a rectangular base acceleration pulse of height A and width T. For $b = 0$, Eq. (22) yields

$$|\ddot{u}(t)|_{\max} = 2A\begin{Bmatrix} \sin\frac{1}{2}\omega_nT, & 0 \le \omega_nT \le \pi \\ 1, & \omega_nT \ge \pi \end{Bmatrix},$$

$$|r(t)|_{\max} = \omega_n^{-2}|\ddot{u}(t)|_{\max}. \qquad (27)$$

Clearly, $|\ddot{u}(t)|_{\max} < A$ only for $\omega_n < \pi/3T$, so the natural period of the oscillator, $T_n = 2\pi/\omega_n$, must be at least six times the pulse width T to achieve acceleration reduction. Over this range, $\sin\frac{1}{2}\omega_nT \approx \frac{1}{2}\omega_nT$, so $|\ddot{u}(t)|_{\max} \approx \omega_nAT$ and $|r(t)|_{\max} \approx \omega_n^{-1}AT$; thus, the price of reduced acceleration is increased rattlespace.

A useful measure of cushion effectiveness is[14]

$$e_c = \frac{V^2}{|\ddot{u}(t)|_{\max}|r(t)|_{\max}}, \qquad (28)$$

where $V = |\int_0^\infty \ddot{w}(t)\,dt|$ is the acceleration impulse. In the preceding example, $e_c \approx 1$; however, if the damper in Fig. 1 is exploited ($b > 0$), performance can be markedly improved. This is shown in Fig. 6, where the damping ratio $\zeta = 0.4$ emerges as optimum, producing $e_c \approx 1.9$. Note that peak relative displacement steadily decreases with increasing ζ, but peak acceleration of the mass grows linearly with ζ for $\zeta \ge 0.5$. Actually, the curves in Fig. 6 are exact only in the limit $T \to 0$ but are quite accurate for $(\omega_nT)^2 \ll 1$. Cushion design is, of course, not limited to linear mass and damper assemblies; many cushions are nonlinear. However, a calculation of e_c is helpful in evaluating almost any cushion.

In some applications, damage or failure is threatened by contact between fragile systems, which is addressed by increasing rattlespace or increasing attachment stiffness; the latter, however, tends to increase peak acceleration. In other applications, the excitation is oscillatory, which could occur when the system of interest is attached to a large, flexible base structure. In this situation, care must be taken to avoid resonance excitation by keeping the fixed-base natural frequencies of the system away from the natural frequencies of the base structure.

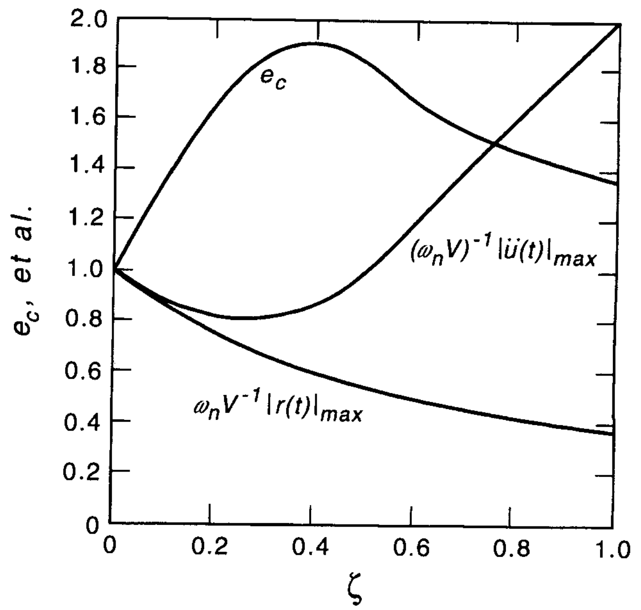


Fig. 6 Effectiveness of a spring–damper cushion as a function of damping ratio.

9 SHOCK SPECIFICATIONS AND TESTING

As might be expected, the most stringent shock specifications are typically associated with military requirements. Representative are the U.S. specifications contained in MIL-STD-810, which addresses such environments as gunfire, pyrotechnic shock and transportation. U.S. Navy specifications for underwater shock environments are contained in MIL-S-901D. International shock standards are contained in ISO 2017-1982 (isolators), ISO 8568-1989 (shock machines), and ISO 9688-1990 (analytical methods).

Shock tests range from a simple package drop test to a series of underwater shock tests conducted against an operating ship. A common shock qualification method involves subjecting a piece of equipment to one or more tests on a shock machine.[15] Such a machine generally consists of a mounting platform that is abruptly set into motion by a hammer, projectile, or quick release from a preloaded condition; or, the motion of a moving platform may suddenly be arrested, as in a drop test. A versatile shock machine may be calibrated to approximate pulses or shock spectra representative of actual environments.

The most common instruments for measuring shock response are accelerometers, velocity meters (or seismometers) and strain gages.[16] The most common accelerometer configuration is that of Fig. 1, in which the spring is a piezoelectric disk in parallel with another (precompressed) spring; the disk generates electric charge in response to the force across it. Because the damping is usually small and no external force is applied to the mass, the equation of motion is (18). For motions with frequency components substantially smaller than ω_n, the force across the piezoelectric disk is $f_r(t) = k_d r(t) \approx -\alpha m \ddot{w}(t)$, where k_d is the spring constant of the disk and α is the ratio of k_d to the sum of the two spring constants. From Eq. (18), this relation is accurate to within 4% for frequency components less than $0.2\omega_n$, where ω_n typically exceeds $2\pi(10^5\text{Hz})$.

Most velocity meters consist of a bar magnet suspended at the axis of a rigidized wire coil through which an electric current flows. The governing equation of motion is again (18), but with ω_n substantially smaller than the frequency components of interest. As the coil moves with acceleration $\ddot{w}(t)$, the lines of magnetic flux inside the coil are cut by the nearly stationary bar magnet, which produces a signal in the coil that is proportional to the relative velocity $\dot{r}(t) \approx -\dot{w}(t)$. Velocity meters are generally accurate over the frequency range 10–1000 Hz; their records can be seismically corrected by means of Eq. (18) to decrease the low-frequency limit.

A modern strain gage consists of a thin metal foil etched to form a long switchback wire that exhibits a change in resistance due to a change in gage length. Thus, it directly measures the average strain over its length by generating, in a dynamic environment, fluctuations in an electric current running through it. When securely bonded to a firm surface, a strain gage produces a highly accurate measurement over a frequency range that is limited by the associated electronic components. Strain gages may be used in the construction of other kinds of transducers for measuring displacement, acceleration, force, and pressure. See Chapter 72 for a more extensive discussion of instrumentation.

10 STRUCTURE–MEDIUM INTERACTION

Many important shock problems involve the interaction between a structure and a surrounding fluid or solid medium; typical examples pertain to airblast, underwater shock, and earthquake environments. The most challenging aspect of these problems is the unboundedness of the surrounding medium. There is some irony in this, because there is usually little interest in the medium per se; the only real interest is in the medium's effect on the structure.

A key strategy in treating structure–medium interaction is *separation* of the total dynamic field in the medium into an *incident-wave* field and a *scattered-wave* field. The incident-wave field is that which exists when the structure is removed and the resulting void is filled with medium. The scattered-wave field is the difference, at every point in the medium at any time, between the total field with the structure in place and the incident-wave field. Determination of the incident-wave field is a separate exercise; its definition is analogous to the specification of $f(x, t)$ in Eq. (2). Determination of the scattered-wave field is difficult in all but the simplest problems.

When the medium is air, the structure may be considered rigid, that is, the scattered-wave field is essentially unaffected by structural motion; exceptional cases involve very flexible structures, such as balloons. When the rigid-body approximation is applicable, the scattered-wave field may be regarded as having been reduced to a *reflected-wave* field. For incident-wave overpressures less than about 35,000 Pa (5 psi), as occur in sonic booms, air may be treated as an acoustic medium, and the fluid response problem reduces to that of rigid-body transient acoustic scattering. Higher overpressures introduce the complexities of nonlinear gas dynamics. In either case, structural response may be determined by applying the sum of the incident-wave and reflected-wave pressures at the surface of the structure.

When the medium is water or soil, however, complete dynamic interaction between the structure and medium occurs: motion of the medium forces the structure to respond, and motion of the structure alters the scattered-wave field in the medium. This is illustrated by the following example: A uniform semi-infinite bar is attached to the mass of the linear oscillator in Fig. 1; a transient incident wave propagates along the bar from right to left, exciting the attached oscillator.

The equations of motion for this example are (1) with $w(t) = 0$ and (2) with $\rho(x) = \rho_0$, $A(x) = A_0$, $E(x) = E_0$, and $f(x, t) = 0$. The boundary conditions for the bar are $u(0, t) = u(t)$ and $A_0\sigma(0, t) = f(t)$, where $\sigma(0, t)$ is the stress at the end of the bar. The initial conditions for the oscillator are $u(0) = \dot{u}(0) = 0$ and those for the bar are $u_S(x, 0) = \dot{u}_S(x, 0) = 0$; $u_S(x, t)$ is the scattered-wave displacement field. The shock-response solution is efficiently sought as follows.

First, one Laplace transforms the simplified Eq. (2) to obtain Eq. (8), the solution to which is $U(x, s) = U_I(x, s) + U_S(x, s)$, where the transformed incident-wave displacement $U_I(x, s) = C_+(s)e^{sx/c}$ and the transformed scattered-wave displacement $U_S(x, s) = C_-(s)e^{-sx/c}$. Also, because $\sigma(x, t) = E_0\epsilon(x, t) = \rho_0 c^2 u'(x, t)$ in the bar, $\Sigma_S(x, s) = \rho_0 c^2 U_S'(x, s)$ is identified as the Laplace transform of the scattered-wave stress field. Second, one uses these expressions for $\Sigma_S(x, s)$ and $U_S(x, s)$ to obtain $\Sigma_S(x, s) = -\rho_0 c s U_S(x, s)$, inverse transforms, and expresses the result at $x = 0$ to obtain the *temporal impedance relation*

$$\sigma_S(0,t) = -\rho_0 c\dot{u}_S(0,t). \tag{29}$$

Third, one introduces this relation and the separation statement $\sigma(x,t) = \sigma_I(x,t) + \sigma_S(x,t)$ into the second boundary condition to obtain $f(t) = A_0[\sigma_I(0,t) - \rho_0 c\dot{u}_S(0,t)]$. Also, the first boundary condition and the separation statement $u(x,t) = u_I(x,t) + u_S(x,t)$ are used to obtain $\dot{u}_S(0,t) = \dot{u}(t) - \dot{u}_I(0,t)$. Finally, using this last equation to eliminate $\dot{u}_S(0,t)$ in the $f(t)$ equation and introducing the result into Eq. (1) with $w(t) = 0$, one obtains the equation of motion for the "medium-encumbered" structure

$$\begin{aligned} m\ddot{u}(t) + (b + \rho_0 cA_0)\dot{u}(t) + ku(t) \\ = A_0[\sigma_I(0,t) + \rho_0 c\dot{u}_I(0,t)]. \end{aligned} \tag{30}$$

For an incident wave prescribed as $u_I(x,t) = \mu_I(t + x/c)$, $\sigma_I(x,t) = \rho_0 c^2 u_I'(x,t) = \rho_0 c\mu_I(t + x/c)$, Eq. (30) is readily solved by analytical or numerical methods, as discussed in Sections 4 and 5.

It is interesting to examine in Eq. (30) the nature of the encumberance imposed by the medium (the bar) on the structure (the oscillator). First, the medium provides acoustic damping through the term $\rho_0 cA_0$; if $\rho_0 cA_0$ is small (an "air bar"), the term $\rho_0 cA_0\dot{u}(t)$ may be neglected, which produces rigid-body scattering as discussed above. The term $\rho_0 cA_0\dot{u}_I(0,t)$ on the right is the reflected-wave force and cannot be neglected; in fact, the equations immediately following Eq. (30) produce, on the right side of Eq. (30), $\sigma_I(0,t) + \rho_0 c\dot{u}_I(0,t) = 2\sigma_I(0,t)$.

11 DOUBLY ASYMPTOTIC APPROXIMATIONS

Because determination of the transient scattered-wave field is so difficult in practical shock problems involving structure–medium interaction, a variety of *temporal impedance approximations* have been proposed. Among these are doubly asymptotic approximations (DAAs),[17] which approach exactness in both the early-time (high-frequency) and late-time (low-frequency) limits. A second-order DAA for a boundary-element model[18] of the wet surface of a structure submerged in an *acoustic fluid* is

$$\ddot{p}_S(t) + c\Gamma_1\dot{p}_S(t) + c^2\Gamma_2\Gamma_1 p_S(t) = \rho c\dddot{u}_S(t) + \rho c^2\Gamma_2\ddot{u}_S(t), \tag{31}$$

where $p_S(t)$ and $u_S(t)$ are, respectively, computational column vectors of nodal scattered-wave pressures and fluid velocities normal to the surface; ρ and c are, respectively, the mass density of and speed of sound in the fluid; and Γ_1 and Γ_2 are constant matrices, each having the dimension of reciprocal length. Equation (31) constitutes a set of ordinary differential equations that are solved simultaneously with the semidiscretized structural equations of motion discussed in Section 5.

If, in Eq. (31), only the first term on the left and the first term on the right are retained, the second-order DAA reduces to $p_S(t) \approx \rho c\dot{u}_S(t)$, which corresponds to the temporal impedance relation obtained in the example of the previous section. This is the *plane-wave approximation*, which is accurate at early time during the structure–medium interaction. If only the last term on the left and the latter term on the right are retained, the second-order DAA reduces to $p_S(t) \approx \rho\Gamma_1^{-1}\ddot{u}_S(t)$. This is the *added-mass approximation*, which characterizes the incompressible flow of fluid at late time. The second-order DAA for an acoustic fluid has been found to yield satisfactory results in analytical studies and shock tests. Analogous DAAs for *elastic and poroelastic media* have also been formulated that reduce to plane-wave approximations at early time and quasi-static approximations at late time.

Temporal impedance approximations have been discussed above as being applied at the surface of the structure. An alternate approach is to semidiscretize a portion of the surrounding medium along with the structure and apply the temporal impedance approximation at the outer surface of that portion. This makes the structural response computations less subject to the inaccuracies of the approximation and permits the treatment of local nonlinearities in the medium and the structure–medium interaction. The price paid, however, is greater computational cost as well as the possible introduction of Gibbs oscillations, as discussed in Sections 4 and 5.

12 CONCLUSION

In recent decades, shock analysis and design have benefited greatly from the rapid development of computer hardware and software. It is now possible, through finite-element and boundary-element semidiscretization, to construct detailed mathematical models of highly complex problems and to generate copious response data from them. Corresponding capabilities exist in the acquisition and processing of experimental data.

To be useful, however, response data must be translated into damage estimates and failure predictions. For example, short-duration exceedance of yield strain within a small volume of reasonably ductile material hardly causes structural failure; but if the degree, duration, and extent of yield exceedance increase sufficiently, failure does occur. Hence, understanding the impact of temporally and spatially localized response peaks on system damage and failure remains the central challenge of shock response analysis and design.

REFERENCES

1. S. Rao, *Mechanical Vibrations*, 2nd ed., Addison Wesley, Reading, MA, 1986, pp. 128, 144.
2. S. Rao, *Mechanical Vibrations*, 2nd ed., Addison Wesley, Reading, MA, 1986, pp. 383, 384.
3. E. Kreyszig, *Advanced Engineering Mathematics*, 7th ed., Wiley, New York, 1993, Chapter 6.
4. S. Rao, *Mechanical Vibrations*, 2nd ed., Addison Wesley, Reading, MA, 1986, pp. 384–387.
5. E. Kreyszig, *Advanced Engineering Mathematics*, 7th ed., Wiley, New York, 1993, pp. 602–603.
6. O. Zienkiewicz and R. Taylor, *The Finite Element Method*, 4th ed., McGraw-Hill, New York, 1989.
7. T. Belytschko and T. Hughes (Eds.), *Computational Methods for Transient Analysis*, Elsevier Science, 1983, Chapters 1 and 2.
8. C. Harris and C. Crede (Eds.), *Shock and Vibration Handbook*, 2nd ed., McGraw-Hill, New York, 1961, Chapter 8.
9. C. Harris and C. Crede (Eds.), *Shock and Vibration Handbook*, 2nd ed., McGraw-Hill, New York, Chapter 23.
10. E. Kreyszig, *Advanced Engineering Mathematics*, 7th ed., Wiley, New York, 1993, Chapter 10.
11. G. O'Hara and P. Cunniff, "The Shock Spectrum Dip Effect," *J. Sound Vib.*, Vol. 103, 1985, pp. 311-321.
12. P. Cunniff and G. O'Hara, "A Procedure for Generating Shock Design Values," *J. Sound Vib.*, Vol. 134, 1989, pp. 155-164.
13. G. Abrahamson and H. Lindberg, "Peak Load-Impulse Characterization of Critical Pulse Loads in Structural Dynamics," *Nucl. Eng. Design*, Vol. 37, 1976, pp. 35–46.
14. C. Harris and C. Crede (Eds.), *Shock and Vibration Handbook*, 2nd ed., McGraw-Hill, New York, 1961, Chapter 31.
15. C. Harris and C. Crede (Eds.), *Shock and Vibration Handbook*, 2nd ed., McGraw-Hill, New York, 1961, Chapter 26.
16. E. Doebelin, *Measurement Systems*, 4th ed., McGraw-Hill, New York, 1990, Chapter 4.
17. T. Geers and P. Zhang, "Doubly Asymptotic Approximations for Submerged Structures with Internal Volumes," *J. Appl. Mech.*, Vol. 61, 1994, pp. 893–906.
18. J. Kane, *Boundary Element Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1994.

68

ACOUSTIC EMISSION

KANJI ONO

1 INTRODUCTION

Acoustic emission (AE) refers to the generation of transient elastic waves due to the rapid release of stored energy from a localized source or sources within a material or structure.[1,2] Such a transient wave is also called acoustic emission and originates at cracks, defective welds, yielding regions, and so forth in metals and at delamination, fiber, and matrix fracture in composite materials.

The generation of AE is a mechanical phenomenon. Acoustic emission from the stress relaxation at the crack tip due to crack propagation is the most representative, but it can originate from a number of mechanisms in various materials, including ceramics, concrete, plastics, and wood. Mechanical deformation and fracture are the primary sources of AE, while phase transformation, corrosion, friction, and magnetic processes among others also give rise to AE. The elastic waves then travel through the structure to a sensor. The manner of wave propagation greatly influences the nature of AE signals at the sensor; for example, when a section size is comparable to the wavelength, the wave travels as a guided wave (Chapter 64). In thick-walled structures, the surface wave is the dominant mode of propagation. Weak electrical signals produced by the sensor are amplified and characterized according to the waveform shape, intensity, frequency spectrum, and other features.[2]

From the detected AE signals, the locations and activity levels of the sources are determined in real time and utilized for the evaluation of the structure or component. Since AE signals can travel over a long distance, the entire structure can be examined in a single test sequence, for example, during preservice or periodic inspection. Acoustic emission can continuously monitor the integrity of a structure in service under favorable conditions. Materials research is another area where AE is useful because of its sensitivity to dynamic microscopic processes. Acoustic emission is a passive test method in that no direct excitation is applied, unlike the ultrasonic test method. However, AE testing requires the application of a certain stimulus to a test piece. In pressure vessels and piping, hydrostatic pressurization is usually employed to activate AE sources. Acoustic emission has become the inspection method of choice for glass-fiber-reinforced composite vessels and piping and more recently for metallic vessels, including industrial gas transport trailers and railroad tank cars.[3]

It is often difficult to establish a mechanical phenomenon at the source of a detected AE signal. This is due to the high sensitivity of AE techniques (since no other means exist to independently observe most dynamic source phenomena) and to the presence of resonance of the structures and sensors.

2 DETECTION

2.1 AE Signals and Source Functions

Two types of AE signals, burst-type (pulselike) and continuous-type (random noiselike) signals, are usually observed. Burst-type emissions arise from distinct events of elastic energy release, such as crack advances, inclusion fracture, fiber fracture, and delamination. A representative example is shown in Fig. 1*a*. It has a sharp rise, followed by an exponential decay. This ringing pattern arises from various resonances in a structure and

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker

Copyright © 1997 John Wiley & Sons, Inc.

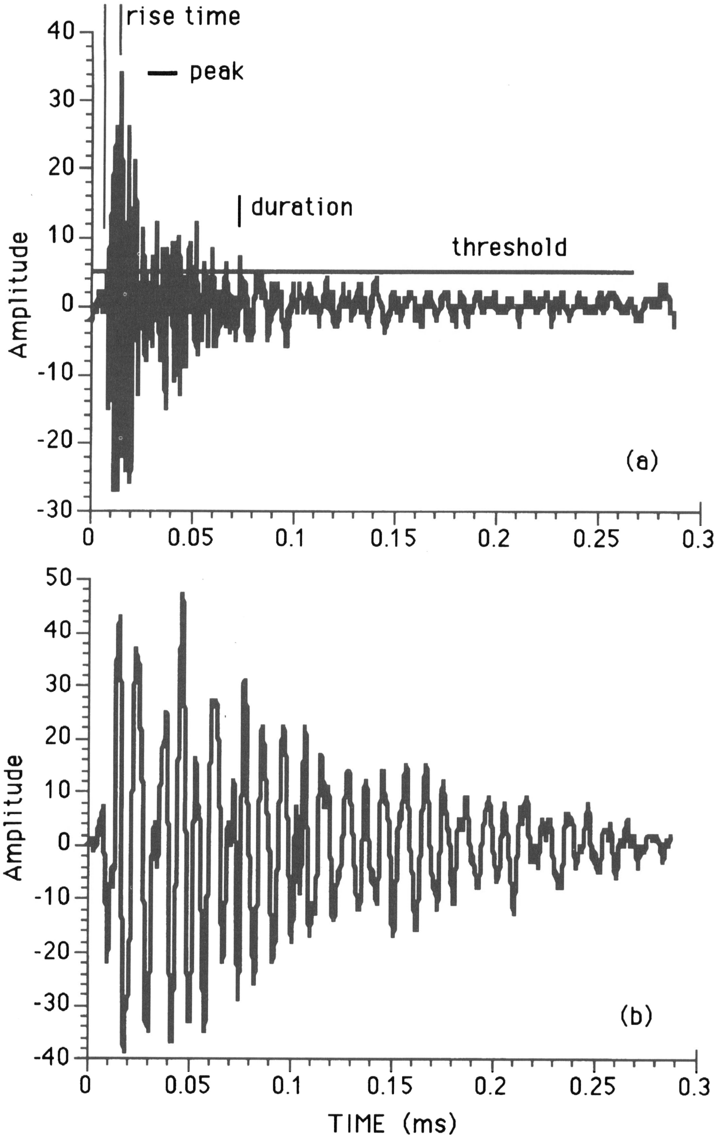


Fig. 1 (*a*) Waveform of a typical (extensional) AE signal from a unidirectional carbon fiber composite. (*b*) Another (flexural) waveform from the same sample. The rise time, duration, and peak amplitude are defined. Amplitude in millivolts after 46 dB amplification. Time axis is 512 μs long.

the resonance of the sensor. Continuous-type emissions are produced by many overlapping events and observed from plastic deformation of metals and from liquid or gas leaks. For typical structural monitoring, ultrasonic frequencies of 30 kHz to 2 MHz are detected. Airborne noise interferes with AE measurements at lower frequencies, while signal attenuation makes the higher frequency range difficult to use. In composite testing, the 30–150 kHz range is common. For applications involving concrete and rocks, a range of several kilohertz to 300 kHz is used, and for geotechnical monitoring the signal frequencies lie below several kilohertz.[1]

Acoustic emission signals contain various information on the AE sources and the propagation paths. The detected signal from a sensor is, in principle, given by convolving the source function (force–time history), the impulse response of the mechanical system to the source (Green's function), and the impulse response of the sensor. For simple geometry, source functions can be recovered by deconvolution when wide-band sensors are employed, and the first motions of signals are discernible. Unfortunately, this is impractical for most applications. Acoustic emission signals detected with a wide-band sensor can be classified according to the generating source types by means of pattern recognition analysis. This scheme primarily depends on the differences in the excitation of various modes of resonance and wave propagation and on their time history.[4]

2.2 Sensors

Most sensors used for AE detection employ a disk of piezoelectric ceramic or crystal and respond to velocity or acceleration normal to the face of a sensor. One type has a broadened frequency response using a backing material behind a transducer element. This construction is essentially identical to that of the ultrasonic sensors (Chapter 52). Another type enhances the sensitivity by using the thickness and/or radial resonance of a transducer element without using a backing material. These are resonant or narrow-band sensors and are employed most often in AE testing because of their high sensitivity. Typical resonant AE sensors have peak velocity sensitivities of a few kilovolts by seconds per metre; these are 20–40 dB more sensitive than those used for conventional ultrasonic testing. The frequency responses of AE sensors are usually calibrated (face-to-face) against a standard using compressional waves at normal incidence. For many applications, however, surface or plate waves are dominant, and their movements normal to the surface are detected, and the size of a sensor affects its responses. Proper sensor selection for frequency ranges or wave modes enhances the sensitivity of AE detection.

High-fidelity capture of signal waveforms is needed for specialized studies of AE sources. Surface displacements are measured using a capacitive sensor or a special sensor with a conical piezoelectric element with matched backing. Laser interferrometers are still not practical due to large size, high cost, and low sensitivity. A sensor calibration system has been constructed using a capacitive sensor in combination with a large steel block and a mechanism to break a fine glass capillary [~0.2 mm outside diameter (OD)]. The breaking of the glass capillary produces a sudden release of force (~20 N within a few tenths of a microsecond) and provides a reproducible source of body waves and surface waves that match well with theoretical predictions. Peak surface displacements of several tenths of a nanometre are generated at 0.1 m from the capillary break. Figure 2 shows the surface wave calibration apparatus, in which a sensor under test is placed at a symmetric position to the capacitive sensor standard.[5] For a working calibration of AE measuring system, breaking of pencil lead (0.3–0.5 mm OD) has been used effectively. This generates ~1 N force drop over a 1-μs period. For concrete, rocks, and geotechnical applications, low-frequency sensors, such as geophones, hydrophones, and accelerometers, are also used. These can have flat frequency responses and some have good directionality.

Given a resonant AE sensor with 3.5 kV·s/m peak sensitivity at 200 kHz, a 1-μV signal corresponds to a 0.6 fm surface displacement over 2 μs. Since a preamplifier (150–300 kHz bandpass) has the input noise of over 0.6 μVrms, this signal cannot be clearly distinguished from noise but is approximately the detection limit. Because AE sensors used in practice have similar sensitivities, the present case may be considered as a reference in dealing

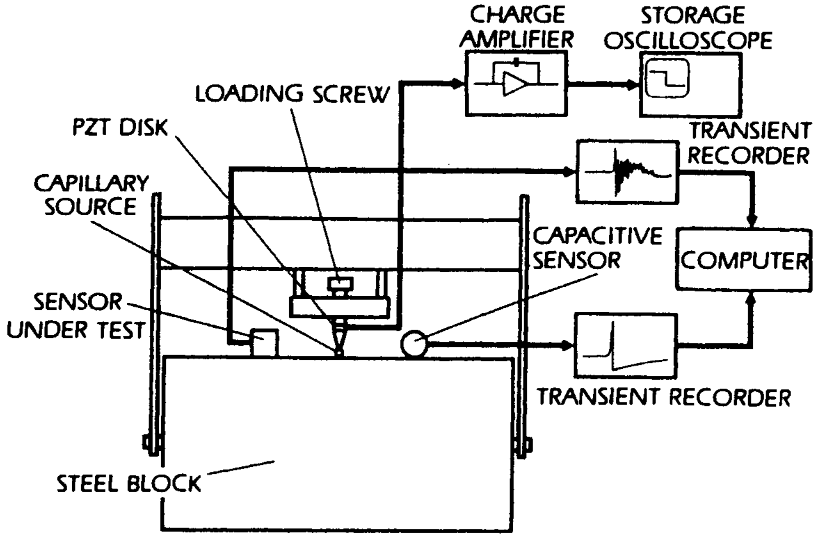


Fig. 2 Schematic diagram of surface pulse apparatus for sensor calibration. Steel block is 90 cm diameter and 43 cm high. Glass capillary source is broken by tightening loading screw and PZT disk measures the breaking force. (Reprinted, by permission, from Ref. 5, p. 88.)

with AE signal amplitude (1 μV at the sensor output is often set at 0 dB).

2.3 AE Signal Characteristics

Characteristics of AE signals in common use[2] are (a) the number (and rates) of burst-type emissions or AE event count (and AE event rates) (in multisensor AE detection systems, the term "hit" is used in place of "event" since a single AE event can reach more than one sensor and counted each time), (b) the averaged signal intensity of continuous-type emissions [e.g., root-mean-square (rms) voltages of amplified signals], (c) the cumulative number or rates of oscillations that cross a prefixed (or noise-compensated) threshold value (called total AE counts or AE count rates), (d) the peak amplitude of burst emissions, (e) the rise time and duration of burst emissions, and (f) the signal strength of burst emission (area of the signal envelope).[3] See Fig. 1*a* for the definition of (d) and (e). In addition, the distributions of peak amplitude, rise time, and duration are also obtained, as well as arrival time differences of burst emissions at different sensors. These signal parameters can be acquired in real time with the advances of electronics. Since the duration of typical burst emissions is 0.1-1 ms, however, at most several thousand signals per second can be distinguished. For higher rates of emissions, average intensity measurements must be utilized. By recording AE signals digitally, the frequency spectrum (assuming the use of wide-band sensors) and shape parameters of the signal waveforms can be determined. Peak intensity and position in the frequency domain, shifts in dominant frequency over time, and rates of rise and decay of the waveform are the features of importance. These features are also used in constructing intelligent classifiers of pattern recognition analysis.[4] Combinations of these and other test parameters, such as applied force or pressure, time under load, ground tilt, and displacement, are used to evaluate the nature of AE sources.

3 SOURCES OF EMISSIONS AND MATERIALS BEHAVIOR

3.1 Plastic Deformation of Metals

Plastic deformation of most structural alloys generates AE that reaches a maximum at or near the yield stress and diminishes with work hardening.[6] Figure 3 shows an example of this behavior of a low-alloy steel.[7] The load–time plot shows a typical discontinuous yielding,

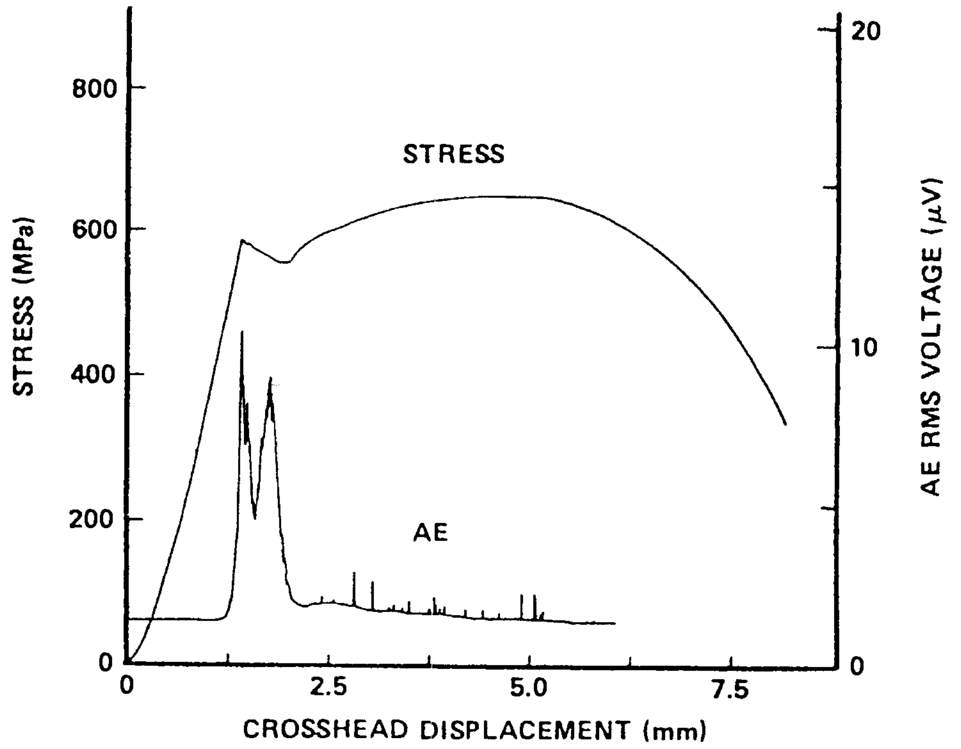


Fig. 3 Load and AE data plotted against the crosshead movement of A533B steel specimen (rolling direction); oil quenched and tempered at 650°C. (Reprinted, by permission, from Ref. 7, p. 21.)

during which AE activities indicated by the rms voltage of AE signals were several times higher than the background noise. This AE upon discontinuous yielding is due to dislocation motion, and the signals are of the continuous type. Purely elastic deformation produces no AE. Once the work hardening begins, the AE activities subside, and the rms voltage is only slightly above the background with occasional spikes due to burst emissions. These burst emissions originate from microcracks or from the fracture and decohesion of nonmetallic inclusions. In some metals, twinning produces burst-type acoustic emission. In metals and alloys showing continuous yielding, AE activities typically start at one fourth to one half of the yield stress and reach a peak at the yield, decreasing gradually with further deformation. At high temperatures where work hardening is low, the AE signal intensity remains strong even at high strains.

Acoustic emission phenomena are irreversible.[6] When a sample is deformed, unloaded, and reloaded, it emits no AE until the previously applied load is exceeded. This is known as Kaiser effect. This behavior is illustrated in Fig. 4*a* and 4*b* where a sample was loaded, unloaded, and reloaded immediately.[7] The Kaiser effect is useful for the determination of a prior loading level, but is not permanent. Extended holding and/or heating before the second loading may reduce the load at which AE starts to be emitted again.

Microstructural variation affects the AE behavior.[6] When a metal or alloy is cold worked, the AE activity is suppressed; often it is eliminated completely. Martensitic and bainitic steels, especially untempered martensites, show very low AE activities as these have high dislocation densities. When the grain size is reduced, AE intensity often increases along with the strength. Other factors may, however, override the grain size effect. In most precipitation-hardened alloys, AE intensity decreases with aging. In Al alloys, the peak AE intensity at yield decreases by more than a factor of 3 from the solution-treated condition to the fully aged condition. When alloys are strengthened by the addition of dislocations or hard precipitates, the mean-free path of mobile dislocations is reduced. When an individual glide motion of the mobile dislocations is restricted, the AE intensity diminishes even though the number of such mobile dislocations increases.

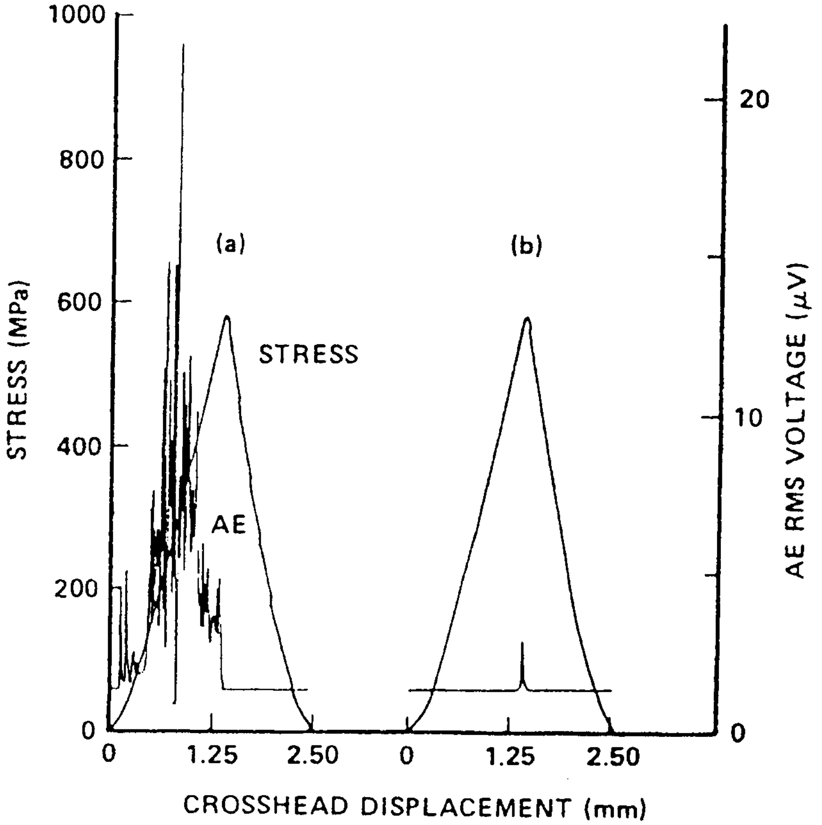


Fig. 4 Stress and the rms voltage of AE signals against time for a steel sample: (*a*) Elastically loaded to 560 MPa and unloaded, (*b*) immediately reloaded after unloading, showing no AE until prior load is reached. (Reprinted, by permission, from Ref. 7, p. 21).

Effects of solid solution impurities and alloying additions vary depending on the strength of solute–dislocation interaction. In alloy systems where the solute pinning of dislocations is weak, the solute additions suppress the AE intensity observed during yielding; cf. dilute Al-Mg, Cu-Al, and Al-Si systems. Quench aging effects occur in Al-Mg alloys, Ni-containing carbon, martensitic steels, and so forth. The AE levels are low in as-quenched states of all these alloys but increase substantially (by as much as 50 times) when solute atoms are allowed to migrate to dislocations, pinning them. When stressed, the pinned dislocations suddenly move and produce dislocation avalanches, generating strong AE activities upon yielding. The strongest AE activities are found in alloys that develop short-range ordering. These include Cu-Zn, Cu-Al, and Ni-Fe. The peak AE levels in these alloys are 10–100 times those of pure metals or usual solid-solution alloys. Dynamic strain aging also produces a series of strong AE peaks. It is found in a number of alloys over a certain range of temperature and strain rate. The AE peaks correspond to load drops. Small load drops are difficult to detect except by AE. Sudden releases of dislocations that had been immobilized are the basic origin of the load drops and bursts of AE activities.

Test temperature alters the AE behavior of materials undergoing plastic deformation. In pure metals and dilute alloys, the peak AE level at yield initially increased by 50–100% with test temperature, but decreased above T/T_m of 0.3–0.4, where T_m is the absolute melting temperature. In normalized steels, the AE level at yield increased fivefold over −150–150°C, while the yield strength decreased by 30%. Acoustic emission from austenitic stainless steels increased 10–20 times from room temperature to 1000 K, where the AE level reached a maximum. The large changes in stainless steels appear to reflect an increase in the stacking fault energy, thus altering the slip mode from planar to nonplanar with increasing test temperature.

Inclusions and second-phase particles are another important origin of AE during deformation. Inclusions are the significant AE sources during tensile loading of steels in the short-transverse direction and that of most high-strength Al alloys. In steels, the decohesion of MnS inclusions is the most important while the fracture of silicide inclusions causes the AE observed in Al alloys. In steels, inclusion-induced AE starts to appear during the elastic loading (see Fig. 4a) and continues into the work-hardening range. These are burst-type emissions and their number is proportional to the inclusion content. The inclusion-induced AE exhibits anisotropic behavior, that is, the highest number of emissions are found in the short-transverse direction and the least in the longitudinal direction.

3.2 Fracture and Crack Growth

During final fracture, most materials and structures produce large, audible sounds. In high-strength materials, strong elastic waves are generated during fracture. Acoustic emission can be detected long before final fracture and can be utilized in preventing catastrophic failures of engineering structures. This application has been and remains to be the primary impetus of the development of AE technology.

Brittle solids under tension or in bending, including ultra-high-strength steels and ceramics, often generate only small number (<100) of AE signals just before final fracture. Subcritical crack growth is minimal, limiting the AE activities. This behavior is unfortunate because these materials are most likely to fail suddenly, and the need to prevent such failures is great. However, all the microfracture mechanisms that are operative in these materials, that is, cleavage, quasi-cleavage, and intergranular fracture, can be detected easily since a large fraction of the AE events have peak amplitudes above 1 mV or 60 dB in reference to 1 μV at the sensor. In some brittle materials, subcritical crack growth (or fracture process zone ahead of the crack tip) is found.[8] During the subcritical crack growth, AE event counts increase in proportion to mth power of stress intensity factor K_I, which is equal to $Y\sigma\sqrt{a}$, where Y is a geometrical factor, σ is stress, and a is the crack length;

$$N_c = A(K_I)^m, \tag{1}$$

where N_c is the cumulative number of AE events and A is a constant. The cumulative amplitude distribution of such AE signals can be approximated by a power-law distribution of the form

$$N_c = B(V_p)^{-n}, \tag{2}$$

where V_p is the peak amplitude and B and n are constants. In low toughness materials (e.g., alumina) showing cleavage and intergranular fracture, $m = 4$ and $n = \sim 0.5$ (see Fig. 5).[8]

When the fracture toughness values increase to 10–100 MPa · $m^{1/2}$, as in high-strength Al and Ti alloys and in high-strength steels, AE signals are produced as the plastic zone ahead of the crack tip expands. Acoustic emission activity becomes significant with many high amplitude emissions (>60 dB) when the crack starts to grow. In this stage, various mechanisms of microfracture may be involved. That is, in addition to those mentioned above, low-energy tear, alternating shear, and microvoid are observed. Acoustic emission event counts increase with stress intensity factor according to Eq. (1) with

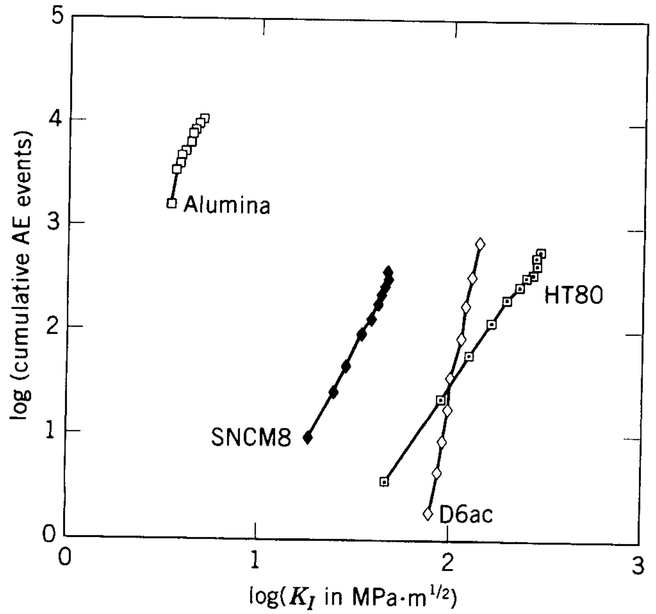


Fig. 5 Log–log plot of AE total events vs. stress intensity factor for alumina and steels. (Data from Refs. 8–11.)

m = 2–20 and the power-law amplitude distribution with n = 0.7–1.5.[9–11] The exponent m often increases as K_I approaches the fracture criticality. Transition in microfracture mechanisms can be reflected more clearly by plotting cumulative AE energy against J_I (cf. Fig. 6).

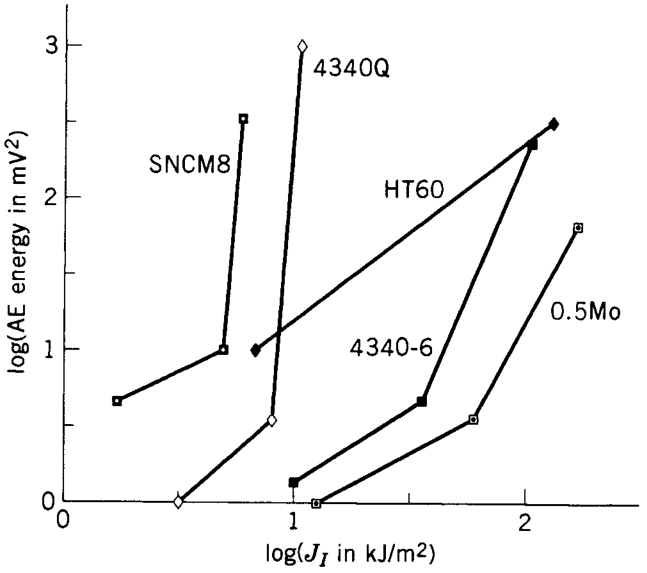


Fig. 6 Log–log plot of cumulative AE energy vs. stress intensity factor for steels. (Data from Refs. 9, 10.) SNCM8 (tempered at 200°C); 4340Q and 4340-6 (4340 steel, as-quenched and quenched and tempered at 600°C); HT60 and 0.5Mo are low-carbon structural steels.

Here, AE energy is defined as the square of peak amplitude, and J_I is given by $(1 - \nu)K_I^2/E$, where E and ν are Young's modulus and Poisson's ratio. The transition is sharp when cleavage microcracks initiate as in SNCM8 and as-quenched 4340 steels but is also visible when fibrous microfracture begins as in 0.5Mo and quenched and tempered (600°C) 4340 steels.

Higher fracture toughness materials fail typically under plane stress conditions by the microvoid coalescence and shear mechanisms with some tearing. In these ductile solids, the expansion of plastic zones ahead of the crack tip initially generates AE, with AE activity reaching a maximum at general yield just like the AE behavior of plastic deformation. This is the weakest among fracture-related AE with the peak amplitude of < 40 dB. Beyond the general yield, stable crack growth processes begin. It is often difficult to detect this by AE as the AE amplitude is still low (< 60 dB) reflecting high microscopic ductility of the materials. In ductile HT60 steel, the transition in AE energy was absent (cf. Fig. 6).

In moderate to high fracture toughness materials, nonmetallic inclusions exert substantial influence on the AE behavior.[6] MnS inclusions in steels and silicate inclusions in Al alloys have strong effects on AE. The decohesion of MnS inclusions emits moderately strong AE (peak amplitude of < 55 dB) from the early part of elastic loading and is the primary source of strong emissions in ductile steels. Plastic flow and fracture of inclusions (especially in Al alloys) also produce AE events. Still, even the maximum AE intensity due to inclusions is less than that from brittle fracture in low toughness materials. This inclusion effect is strongest in samples stressed in the short-transverse direction. The total counts at the maximum load for the short-transverse samples are 20–50 times that of longitudinal samples of the same steel.

Fatigue leads to eventual fracture in a number of engineering structures under repeated loading.[1] Acoustic emission accompanies fatigue crack initiation and subsequent growth, showing rapid rise just before final fracture. Often, the detection of AE precedes an optical observation of a fatigue crack. In cyclically loaded structures, similar trends have been found. Once a crack develops, different types of AE are emitted as a function of loading cycle phase.

(a) Near a peak tensile load, AE due to crack growth and inclusion fracture is observed.

(b) As the load is reduced to zero, crack closure noise is detected as AE; that is AE originates from crack face fretting and crushing of oxide particles.

(c) As the load increases, crack faces that had stuck together separate, producing AE.

Peak-load AE contributes less than 10% of the total AE activity in some cases, but more when high stress ranges are used. Initially, AE events are detected over a wide range of loading but tend to concentrate at the maximum load toward the end of fatigue loading. For the detection of fatigue cracking, frictional AE due to the fretting and crack face separation are important. In addition, particles of oxides and various corrosion products often form between the crack faces. Their subsequent fracture also contributes to the AE due to fatigue. Under high crack growth (plane stress) conditions, more AE is observed, but AE activity per unit crack area is reduced.

3.3 Fiber Fracture and Delamination of Composite Materials

Most composite materials in current use are reinforced by glass fibers, which take various forms, such as short random fibers, mat, woven rovings, and continuous fibers. Aerospace composites use higher modulus carbon and aramid fibers. The matrix materials include thermosetting plastics (polyester and epoxy) and thermoplastics (nylon and polysulfone). These resin–matrix composites are difficult to inspect with conventional nondestructive test methods and AE has been used widely.[12] Main sources of emissions are fiber fracture, delamination (matrix cracks between reinforcement layers), splitting (matrix cracks parallel to fibers), and transverse matrix cracks.

When unidirectionally reinforced composites are stressed, fiber fracture and matrix cracks contribute to AE.[3,4,12] Glass fibers generate high amplitude emissions (>70 dB in reference to 1 μV at the sensor) just prior to composite fracture (>90% of the fracture load). Carbon fibers start to fracture above 50% of the fracture load of the composites and produce low-amplitude (30–60 dB) emissions, reflecting their smaller diameters and lower fracture strain. The rate of emissions increases rapidly as the final fracture load is approached, similar to the behavior of brittle solids. Near the final fracture, high-amplitude, long-duration signals from splitting are also observed. Delamination occurs when reinforcement layers (called plies) have different fiber orientations or notches. Acoustic emission signals from delamination are the strongest (50–130 dB peak amplitude) and have the longest duration (0.1–10 ms).

Most composite structures are fabricated with multiple plies with various combinations of reinforcement orientations. When they are stressed, complex stress patterns develop between reinforcing fibers and plies and copious emissions are generated, mainly from the transverse matrix cracks and delaminations. Discrimination of individual AE signals relies on peak amplitude, rise time, and duration. Indication of damages in glass fiber composites include high-amplitude emissions (>70 dB), long-duration signals, and emissions during load hold periods. However, signal discrimination is often difficult because of high rates of emissions and strong signal attenuation.[3]

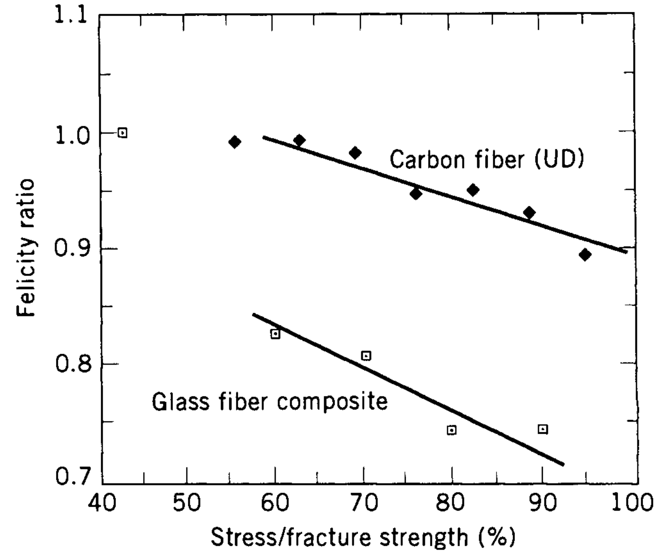


Fig. 7 Felicity ratios of a glass fiber composite panel and a unidirectional carbon fiber composite sample. (Glass fiber data from Ref. 13.)

When a composite contains damages, the Kaiser effect is no longer observed. The Felicity ratio is defined as a ratio of the load, at which AE is observed upon reloading, to the maximum prior load. The Felicity ratio is 1.0 when the composite is sound and decreases to lower values (0.6–0.8) after severe loading. As shown in Fig. 7, the Felicity ratio drops to 0.8 by stressing above 70% of the fracture load in glass-fiber-reinforced composite.[13] On the other hand, a unidirectional carbon fiber composite retains high Felicity ratios (>0.9) up to 95% of the fracture load.

3.4 Other Emission Sources

In a corrosive environment, AE is observed from hydrogen evolution, pitting, and exfoliation, but anodic metal dissolution produces no AE. When materials are subjected to stress corrosion or hydrogen environments, they produce AE by a number of different mechanisms, including film breakage, hydrogen-induced cracking, inclusion and particle fracture at the crack tip, and plastic deformation.[14] The detection of hydrogen embrittlement is one of the early applications of AE techniques. During crack propagation due to hydrogen embrittlement, high-amplitude emissions are observed and can be distinguished easily from active path corrosion process. The latter generates only a small number of low-

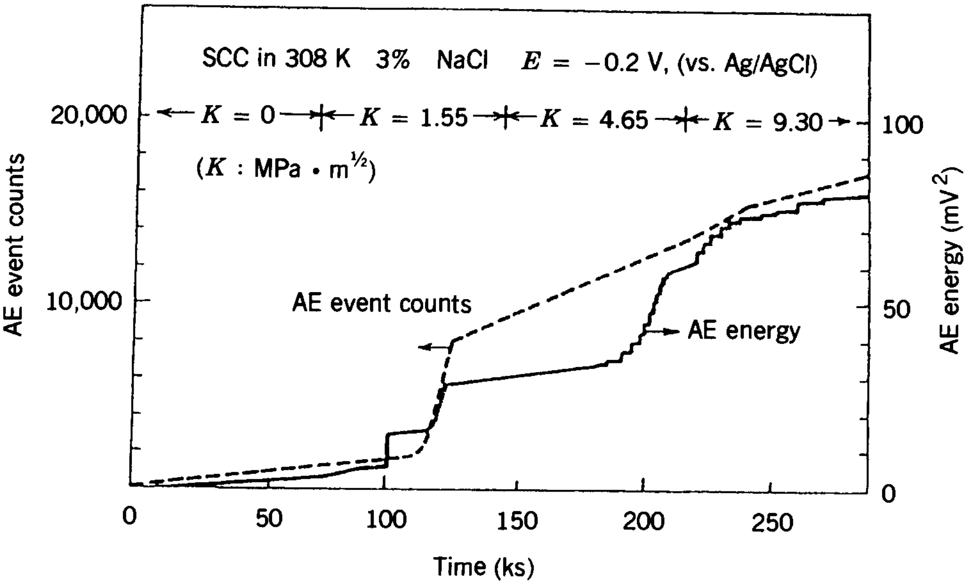


Fig. 8 Cumulative AE event counts and AE energy during stress corrosion cracking test under four different stress intensity factor levels; 304 stainless steel was exposed to a saline solution. (Reprinted, by permission, from Ref. 14, p. 75.)

amplitude emissions. The crack growth rate for stress corrosion cracking shows the three-stage behavior with respect to the applied K_I; that is, an initial increase above a threshold K_I, followed by a region with a nearly constant rate and the final rise near the critical K_I. The AE activities show a similar behavior. Cumulative AE event counts and AE energy are plotted against time in Fig. 8, where 304 stainless steel was exposed to a saline solution with varying levels of stress intensity factor K_I. When $K_I = 0$, AE is due to hydrogen evolution and has low amplitude only. For K_I above 1.5 MPa · $m^{1/2}$, microcracking due to stress corrosion is detected with high-amplitude AE signals.

Acoustic emission monitoring of welding processes can locate weld cracks and slag inclusions and find the depth of weld penetration and mistracking. Expulsion and lack of weld in spot-welding can also be detected. These in-process monitoring methods may avoid the electrical welding noise by using guard sensors and gating circuits synchronized to welding steps. Evaluation of weldments has been the primary goal of AE monitoring of welded structures, such as pressure vessels, tanks, and highway bridges. Defective welds can be located by AE and further evaluated by other nondestructive test methods.

Monitoring of various machining or metal removal processes detects tool wear, fool fracture, and cutting conditions. These are based on AE due to friction between tool and chips or a workpiece, abnormal signals from chipped tools, and plastic deformation and fracture of chips. This application is increasing its importance with automated manufacturing processes.

Sudden movements of magnetic domain walls in ferromagnetic materials produces AE. This is called magneto-AE, and its origin is related to magnetostriction. The intensity and waveform of magneto-AE depend on applied or residual stress, chemical composition, heat treatment, prior cold working, and test temperature. This is an acoustic equivalent of the Barkhausen effect and can be used to evaluate residual stress and microstructural changes of a ferromagnetic material.

The following are other important AE sources of AE: in metals, martensitic transformation, solidification (hot cracking), and oxidation (oxide cracking and spalling); gas and fluid leakage; corona discharges in power transformers; and friction and rubbing noise from rotating machinery.

4 SOURCE LOCATION

An important function of AE testing is to locate sources in a structure.[1] Acoustic emission testing uses the structure itself to discover discontinuities through the generation and propagation of AE signals, which can be continuous (e.g., fluid leaks and plastic deformation of metals) or burst emission (e.g., crack propagation and delamination of composites). Two sensors are used to locate the source in a linear structure such as a pipe and a tube (linear location). More sensors are needed when the

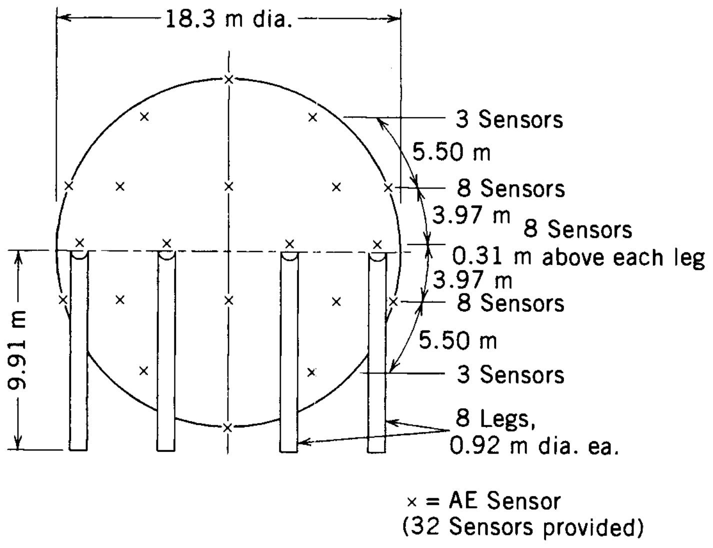


Fig. 9 Sensor placements on a spherical pressure vessel. (Reprinted from Ref. 15, p. 436.)

linear structure becomes longer and when three-dimensional structures are examined. Figure 9 shows typical sensor placements on a spherical pressure vessel.[15] In most applications, the structures can be regarded as a two-dimensional shell structure and the surfaces can be laid out on a plane. For example, a spherical tank can be represented by an icosahedron and by 20 triangles on a planar display.

Two general approaches are used in source location; that is, zone location and discrete source location. In the zone location method, a source is presumed to exist within the zone that belongs to the sensor receiving the strongest AE signal (typically in peak amplitude) or receiving the first hit signal. This method is suited for locating leaks (continuous emissions) and for sources in a highly attenuating medium (fiber composites and metals coated with viscous insulation). The method can be refined by determining which sensor receives the second strongest signal or the second hit signal. In this method, the zone around each sensor is subdivided, and the number of the subdivision is equal to that of surrounding sensors. When the amplitudes of the received signals are measured, further improvements in source location accuracy are possible through an interpolation technique by taking the attenuation into account.

In discrete source location, the full spatial coordinates of an AE source are defined by measuring the differences in arrival times of AE signals at multiple sensors. The AE signals must be burst type and strong enough to reach three or more sensors. In attenuating media, the sensor spacing has to be sufficiently small. Generally, the wave velocity is assumed constant and triangulation techniques are used. A pair of sensors define an arrival time difference and a hyperbola passing through a source. The location of the source can be obtained as the intersection of two such hyperbolae. Hundreds of sources per second can be located by typical AE instruments. For this method to be practical, the attenuation through the structure cannot be high to avoid the use of excessive sensors and signal processing channels. Emission rates must be moderate so that the arrival time differences of a single event can be detected without interference of the next event.

In order to evaluate the severity of an AE source, the intensity and activity of the source and proximity to neighboring sources are determined. This requires the collection of additional data on the AE sources and typically posttest analysis. Such data include the presence of high-amplitude emissions, AE activities during load hold, those during stress increases, and Felicity ratios during load cycling. Statistical analysis (e.g., amplitude distribution) of AE signals received at individual sensors is also useful.

5 STRUCTURAL INTEGRITY MONITORING

5.1 Pressure Vessels, Tanks, and Piping: Fiber-Reinforced Composite

No satisfactory test for determining the structural adequacy of fiber-reinforced plastic equipment existed until AE testing of fiber-reinforced plastic vessels and piping has been developed and standardized.[3] Test vessels and piping are typically pressurized to 110% of the maximum allowable working pressure to locate substantial flaws. Using the zone location method, AE activities of flaws within each zone are detected and the zone represents the approximate position of these flaws. Sensors are positioned to detect structural flaws at critical sections of the test vessel (i.e., high-stress areas, geometrical discontinuities, nozzles, and manways). Pressurization in AE testing of a vessel proceeds in steps with pressure hold periods with depressure increments. A test is terminated whenever a rapid increase in AE activities indicates an impending failure. Detected flaws are graded using several criteria, including emissions during pressure hold periods, Felicity ratio, historic index (average signal strength of recent events divided by that of all events), severity (average signal strength of 10 largest events), and high-amplitude events. Emissions during hold indicate continuing permanent damage and a lack of structural integrity. For in-service vessels, Felicity ratio criterion (when it is less than 0.95) is an important measure of previous damage. Historic index is a sensitive measure to detect a sudden increase in signal strength. Large severity values result from delamination, adhesive bond failure, and crack growth. High-amplitude events indicate structural (glass-fiber) damages, especially in new vessels.

5.2 Pressure Vessels, Tanks, and Piping: Metals

Acoustic emission testing of metallic tanks and pressure vessels is conducted during pressurization. The tanks and vessels are pressurized following applicable code specifications. The maximum pressure is typically up to 110% of the operating pressure. Sources of emissions are usually crack growth, yielding, corrosion effects (cracking or spalling of corrosion product, local yielding of thinned section), stress corrosion cracking, and embrittlement. The most likely locations of flaws are at various weldments.[15] When metals are ductile, AE activities are low and AE test results should be evaluated carefully. Once the metals are embrittled by environment or at low temperatures, even early stages of stressing activate AE sources. In fact, acoustic emission is the best method for detecting hydrogen-induced cracking and stress corrosion cracking. Highway gas-trailer tube, chemical, and petrochemical vessels are monitored with AE during and between service without emptying the vessels, thus minimizing the cost of testing. According to the current industrial practices, the located flaws are usually confirmed by other nondestructive test methods. Increasingly, however, AE testing alone is used to evaluate tanks and pressure vessels. For certain classes of metallic vessels, the use of AE examination alone can be used to satisfy code-mandated inspection requirements in lieu of ultrasonics or radiography.

After AE testing indicates the locations of active AE sources, these are graded according to criteria similar to those for composite vessels. These provide real-time indications of defective areas of the pressure vessel being tested and prevent catastrophic failure of the vessel. The identified defective areas are then inspected using other test methods. Acoustic emission testing is applied on various types of vessels, including ammonia spheres and tanks, hydroformer reactors, and catalytic cracking units.

5.3 Aerospace

The first application of AE testing was to verify the structural integrity of filament-wound Polaris rocket motorcases in 1961.[16] Hydrostatic testing of rocket motorcases was instrumented using accelerometers, audio recording, and sound analysis equipment. Crack initiation and propagation were detected prior to catastrophic failure. Later, burst pressure was successfully predicted based on AE data during proof testing. The composite rocket motorcases are still successfully inspected using three (or four) pressurization cycles. The first cycle is to find leaks at a low pressure. The second goes to 100% (or 80%) of the mean expected operating pressure, followed by the third to 80% level. The AE behavior during the third cycle is used to examine the Kaiser effect and is the determining factor in the evaluation of the rocket motorcase integrity.

Acoustic emission testing is as yet unable to monitor the integrity of aircraft structures in flight because of substantial noise generated from the structures that are joined by numerous fasteners and from engines. For specific components, however, AE can detect the presence of cracks and incipient fracture. For surveillance of structural and fatigue damages of aircraft components during limit load and fatigue testing, AE has demonstrated its utility. All significant cracks were located in the main wing-spar of a Mirage jet-fighter during full-scale fatigue testing.[17] Some of the cracks emanated from bolt holes amid a wide range of spurious AE sources. Acoustic emission indications of damage severity vary depending on the particular component, but include load hold emissions, Felicity ratio, AE event rates, and high-amplitude emissions. In structural tests, wave propagation characteristics and the nature of crack-related (crack growth and crack face fretting) sources are also essential in evaluating AE observations.

5.4 Geological and Civil Engineering Structures

Acoustic emission from rocks in underground mines was first detected in the 1930s by Obert and Duvall and has been called microseismism, as it is a microscopic earthquake.[18] Acoustic emission monitoring has been studied to promote mine safety by detecting incipient rock bursts and gas outbursts and by estimating the areas of stress concentration ahead of mining regions. For this purpose many AE sensors are placed in and around mine tunnels (some from holes dug into the ground) and monitor low-frequency (below several kilohertz) AE signals. A mining operation itself can be a stimulus as underground rocks have numerous preexisting flaws and AE under geological pressure. Drilling exploratory holes ahead of a mining region can also activate the areas of high local stresses. Multichannel AE sensing has also been applied to the evaluation of hydrofracturing in geothermal wells. The stability of underground storage caverns is another area of AE applications.

By using the Kaiser effect, underground stresses may be estimated.[19] Compressive stresses are applied on a sample bored out from underground. When the previously existing stress is exceeded, AE activities increase and the preexisting stress can be estimated. Since a high compressive pressure always exists, the directionality of stresses must be taken into account. Loading methods and relaxation after coring may have effects on the results, however.

Acoustic emission techniques have been applied to monitoring the stability of highway slopes, tunnels, and landslide-prone areas. Acoustic emission measurements utilize waveguides that are driven into the ground. By mapping high AE activity areas, potential instability regions can be predicted. Concrete with steel reinforcement is another material used widely in civil engineering structures.[20] Under bending, compressive or tensile loading, AE activities increase as the fracture load is approached. Bending and shear modes of fracture can clearly be distinguished via AE. Because of the large size of concrete structures and high attenuation rates of AE signals in concrete, suitable AE techniques for structural integrity monitoring are yet to be perfected.

6 SUMMARY

Acoustic emission is the generation of transient elastic waves due to the rapid release of energy from a localized source or sources within a material or structure, originating from mechanical deformation and fracture as well as from phase transformation, corrosion, friction, and magnetic processes. From AE signals detected by sensors attached to the structure, the locations and activity levels of the sources are evaluated. Acoustic emission has become an indispensable and effective nondestructive inspection method and a valuable tool of materials research, delving into dynamic microscopic processes of materials. A survey of recent AE literature is available.[21]

REFERENCES

1. R. K. Miller and P. McIntire (Eds.), *Nondestructive Testing Handbook*, 2nd ed., Vol. 5, *Acoustic Emission Testing*, American Society for Nondestructive Testing, Columbus, OH, 1987.
2. A. G. Beattie, "Acoustic Emission, Principles and Instrumentation," *J. Acoust. Emission*, Vol. 2, 1983, pp. 95–128.
3. T. J. Fowler, "Chemical Industry Applications of Acoustic Emission," *Mat. Eval.*, Vol. 50, 1992, pp. 875–882.
4. K. Ono and Q. Huang, "Pattern Recognition Analysis of Acoustic Emission Signals," in T. Kishi et al. (Eds.), *Progress in Acoustic Emission VII*, Japan Soc. Nondestructive Inspection, Tokyo, 1994, pp. 69–78.
5. F. R. Breckenridge, "Acoustic Emission Transducer Calibration by Means of the Seismic Surface Pulse," *J. Acoust. Emission*, Vol. 1, No. 2, 1982, pp. 87–94.
6. C. R. Heiple and S. H. Carpenter, "Acoustic Emission Produced by Deformation of Metals and Alloys—A Review, Part I and II," *J. Acoust. Emission*, Vol. 6, 1987, pp. 177–204, 215–237.
7. I. Roman, H. B. Teoh, and K. Ono, "Thermal Restoration of Burst Emissions in A533B Steel," *J. Acoust. Emission*, Vol. 3, No. 1, 1984, pp. 19–26.
8. S. Wakayama, T. Kishi, and S. Kohara, "Microfracture Analysis in Al_2O_3 Evaluated by AE Source Characterization," in K. Yamaguchi et al. (Eds.), *Progress in Acoustic Emission III*, Japan Soc. NDI, Tokyo, 1986, pp. 653–660.
9. K. Kuribayashi, et al., "Fracture Toughness and AE of Structural Low Alloy Steel (SNCM8)," *Hihakai Kensa* (*J. Nondestructive Testing*), Vol. 30, 1981, pp. 842–846.
10. H. Takahashi, M. A. Khan, M. Kikuchi, and M. Suzuki, "Acoustic Emission Crack Monitoring in Fracture-Toughness Tests for AISI 4340 and SA533B Steels," *Exp. Mech.*, Vol. 21, 1981, pp. 89–99.
11. Y. Nakamura, C. L. Veach, and B. O. McCauley, "Amplitude Distribution of Acoustic Emission Signals," in *Acoustic Emission*, ASTM STP 505, Amer. Soc. Testing and Materials, Philadelphia, 1972, pp. 164–186.
12. J. Summerscales, "NDT of Advanced Composites—An Overview of the Possibilities," *Br. J. Non-Destructive Testing*, Vol. 32, No. 11, 1990, pp. 568–77.
13. T. J. Fowler and E. Gray, "Development of an Acoustic Emission Test for FRP Equipment," Preprint 3583, American Soc. of Civil Engineers, New York, 1979.
14. S. Yuyama, T. Kishi, and Y. Hisamatsu, "AE Analysis during Corrosion, Stress Corrosion Cracking and Corrosion Fatigue Processes," *J. Acoust. Emission*, Vol. 2, 1983, pp. 71–93.
15. T. J. Fowler, "Acoustic Emission Testing of Chemical Industry Vessels," in M. Onoe et al. (Eds.), *Progress in Acoustic Emission II*, Japan Soc. Nondestructive Inspection, Tokyo, 1984, pp. 421–449.
16. A. T. Green, C. S. Lockman, and H. K. Haines, "Acoustic Verification of Structural Integrity of Polaris Chambers," *Mod. Plastic*, Vol. 41, No. 11, 1964, pp. 137–139.

17. C. M. Scala, J. F. McCardle, and S. J. Bowles, "Acoustic Emission Monitoring of a Fatigue Test of an F/A-18 Bulkhead," *J. Acoust. Emission*, Vol. 10, Nos. 3/4, 1991/92, pp. 49–60.
18. H. R. Hardy, Jr. (Ed.), *Proc. the Fifth Conference on AE/Microseismic Activity in Geologic Structures,* Trans. Tech. Publ., Clausthal-Zellerfeld, Germany, 1991.
19. T. Kanagawa, M. Hayashi, and H. Nakasa, "Estimation of Spatial Geo-stress Components in Rock Samples Using the Kaiser Effect of Acoustic Emission," *Proc. Japan Soc. Civil Eng.*, No. 258, 1977, pp. 63–75.
20. M. Ohtsu, "Acoustic Emission Characteristics in Concrete and Diagnostic Applications, *J. Acoust. Emission*, Vol. 6, No. 2, 1987, pp. 99–108.
21. K. Ono, "Trends of Recent Acoustic Emission Literature," *J. Acoust. Emission*, Vol. 12, Nos. 3/4, 1994, pp. 177–198.

69

SEISMIC EXPLORATION: ACOUSTIC IMAGING FOR GEOLOGICAL INFORMATION

A. J. BERKHOUT

1 INTRODUCTION

Exploration seismology is based on the analysis of acoustic waves reflected from different rock layers in the earth's subsurface. (Here *acoustic* is used in the elastodynamic sense.) Impulsive acoustic energy, emitted into the subsurface, encounters discontinuities between the layers and is partially reflected back to the surface. The returning reflections are detected and their strengths and arrival times are recorded (*seismic acquisition*). After the recorded data are processed (*seismic processing*), geophysicists and geologists derive information from the result concerning the position, thickness, rock type, and porefill of each layer (*seismic interpretation*).

Most of the recent advances in seismic exploration techniques have been made possible by new digital technology. This holds true in all three seismic disciplines: acquisition, processing, and interpretation. The ability to record and store very large volumes of seismic data, the ability to reduce noise on the recordings and then improve the resolution properties, the ability to manipulate vast quantities of data with advanced imaging algorithms to arrive at complex structural interpretations, and finally the ability to apply multidisciplinary inversion techniques to arrive at the rock properties of potential reservoirs have all been made possible by improved digital technology. And as seismic exploration techniques continue to expand, new systems such as data robots, massive parallel processors, fast workstations with large internal memories, and advanced graphics will be taken for granted as everyday equipment in the exploration world.

The chapter starts with a discussion on the close relationship between acoustic imaging and seismic exploration. This discussion is followed by an explanation of the principles of seismic exploration in terms of ray theory. Next a review is given of the seismic forward model in terms of wave-theory-based operators. This model is used to explain modern seismic processing in terms of *preprocessing* of the raw measurements, image *formation* (*seismic migration*) and image *characterization* (*lithologic inversion*).

2 ACOUSTIC IMAGING AND SEISMIC EXPLORATION

Acoustic reflection techniques aim at high-resolution images of objects under investigation. Figure 1 shows three important commercial applications, together with the corresponding media. Later in this chapter we will show that propagation properties can be quantified by propagation matrices $\boldsymbol{W}^+$, $\boldsymbol{W}^-$ and reflection properties can be quantified by reflection matrices $\boldsymbol{R}^+$, and $\boldsymbol{R}^-$.)

In *medical diagnostics* the objective is to collect information on the tissues inside the human body. In *seismic exploration* the geologic layers below the earth's surface are of interest. In *ultrasonic inspection* defects in construction materials need to be detected. Although the instrumentation in those applications is significantly different, they have the underlying acoustical principles in common.

In acoustic reflection techniques, the medium under investigation is illuminated with acoustic *waves (insonification)*. The induced acoustic wave field (source wave field) propagates from a data acquisition surface into the medium, reflects at the inhomogeneities, propagates back

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker
Copyright © 1997 John Wiley & Sons, Inc.

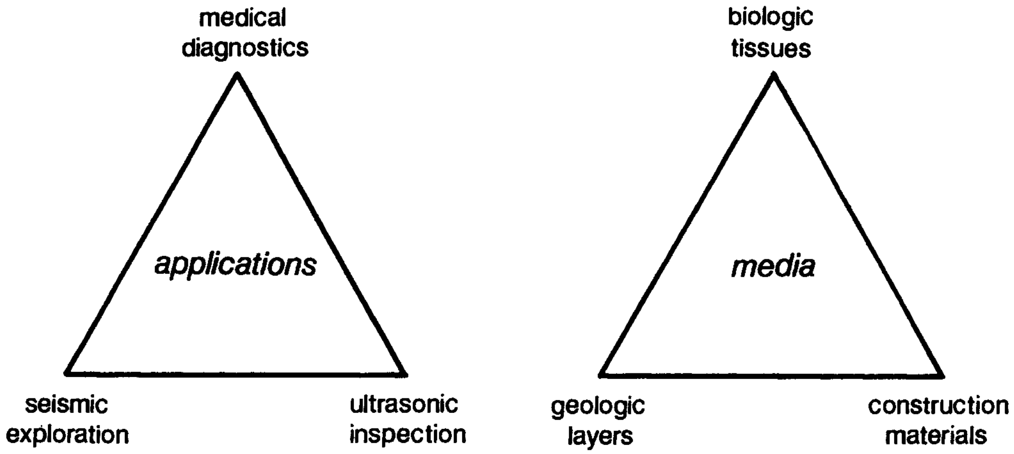

medical diagnostics
applications
seismic exploration
ultrasonic inspection
biologic tissues
media
geologic layers
construction materials

Fig. 1 Three important commercial applications of acoustic reflection techniques, together with the corresponding media.

to the surface (reflection response), and is measured by a range of acoustic detectors. Information about the structural and material properties of the interior is derived from the recorded detector signals. For many practical problems it is necessary to illuminate the medium from different locations at the data acquisition surface (different source positions), and the registered reflection responses of the different acoustical experiments may be optimally combined (processing). Later in this chapter we will see that reflection responses from different experiments can be conveniently represented by measurement matrix $\boldsymbol{P}$.

The properties of the detected waves are determined by two properties of the medium. The first are the *propagation* (or transmission) properties, which primarily depend on the *global* acoustic properties of the medium such as average velocity and average absorption; if applicable, the average values should be supplemented by gradients (trend model). The second are the *reflectivity* properties, which primarily depend on the *local* acoustic properties of the medium such as detailed variations in the modulus of elasticity (Fig. 2). The importance of the trend model concept can be readily appreciated by realizing that the relation between the reflection arrival times and the position of the corresponding reflection points are determined by the velocity trend model of the object. Note that in the inversion literature the trend model is generally referred to as the *background medium.*

In medical diagnostics the background is taken to be a homogeneous half-space; in ultrasonic inspection the background is represented by a *bounded* homogeneous medium; in seismic exploration the background is a half-space that consists of a number of macrolayers, where each macrolayer is characterized by its own trend parameters. In seismic exploration the background is generally referred to as the macromodel.

In principle, acoustic imaging aims at determination of both trend and detail. Particularly in seismic exploration, the estimation of the macromodel is of utmost importance before imaging for the detail can start. However, in medical as well as inspection applications an

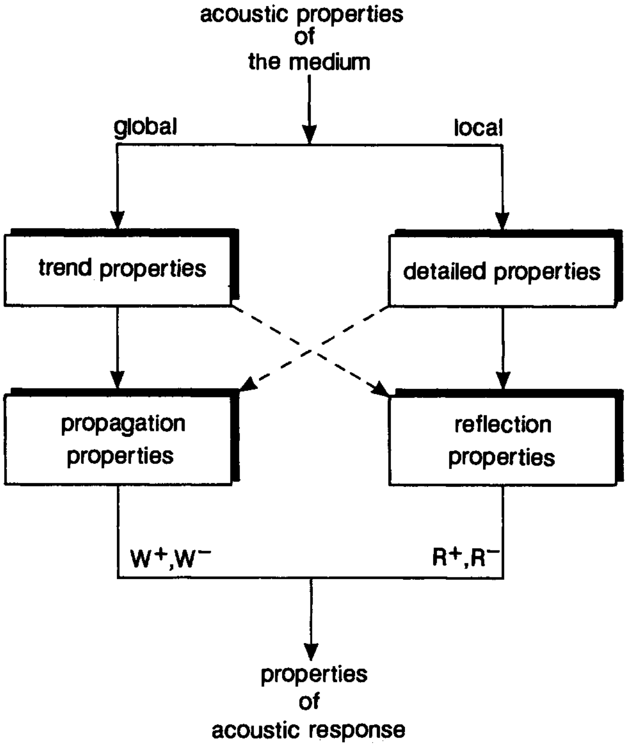

acoustic properties of the medium
global
local
trend properties
detailed properties
propagation properties
reflection properties
W+,W−
R+,R−
properties of acoustic response

Fig. 2 Relatively slow trend properties of the medium determine the propagation (transmission) properties of the response, i.e., primarily the travel time information. The relatively fast detailed properties of the medium determine the reflection (backscatter) properties of the response, i.e., primarily the amplitude information.

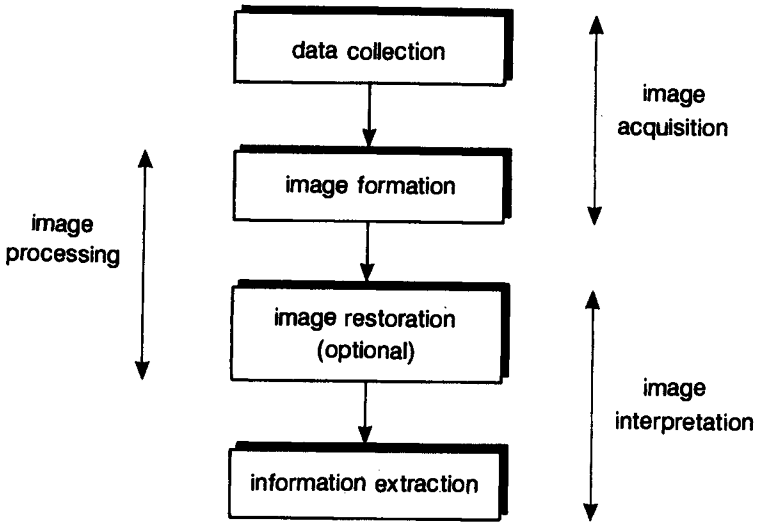


Fig. 3 Basic processes in reflection imaging.

estimate of the trend parameters of the medium is already available but local variations are unknown; the objective of acoustic imaging techniques in those applications is to obtain information on the local variations only.

Figure 3 shows the basic processes in acoustic reflection techniques. In its most general form reflection data are generated and digitally recorded (data collection) and optionally preprocessed. Next, the preprocessed data is used to form the actual image numerically (image formation). However, in *focused* experiments image formation is part of the data collection and images are directly recorded. Since data collection and image formation are never perfect in practice, the imperfections are optionally reduced by *image restoration*. The result is used as input to the final information extraction process (interpretation). Note that in the imaging literature numerical image formation is often referred to as *aperture synthesis*. Note also that in seismic exploration numerical image formation is always referred to as seismic migration.

In acoustical reflection imaging one has to deal with two fundamental problems:

1. Optimization of the signal-to-noise ratio
2. Optimization of the resolution

Noise is not only determined by statistical background signals but also by undesired wave types and multiple scattering.

The resolution along the time axis (temporal resolution) is determined by the range of frequencies that are present in each individual echo (temporal bandwidth). A large temporal resolution requires a large temporal bandwidth. The spatial resolution is determined by the size of the involved wavelengths together with the range of immergence and emergence angles that are present in the recorded echoes from each individual inhomogeneity (spatial bandwidth). Small wavelengths and a large range of angles are required for images with a high spatial resolution. Note that a large *temporal* resolution favors a large *spatial* resolution.[1]

Imperfect image formation, for example, due to the use of wrong macrovelocity information or limited focusing instrumentation, will decrease the resolution more than could be expected from the limited temporal and spatial bandwidth. It is the objective of the image restoration process to identify and repair the damage to the resolution (*deblurring*), taking the presence of noise into account. In those cases where *additional* information of the object can be specified, image restoration may yield results even beyond the resolution specified by the available temporal and spatial bandwidth.

In practical situations the signal-to-noise ratio decreases with increasing depth due to attenuation (penetration problem). Similarly, resolution decreases with increasing depth due to the relatively large attenuation for high frequencies (small wavelengths). In addition, the aforementioned range of immergence and emergence angles generally decrease with increasing depth, causing an extra problem for *lateral* resolution. Figure 4 provides a view of the applied frequencies in the different applications. Every frequency range is a compromise between resolution and penetration: By shifting toward the high frequencies, the resolution increases but the penetration decreases because of increased attenuation. Note particularly the large difference in utilized frequencies, as implemented in low-frequency seismic exploration and high-frequency ultrasonic inspection. Deep seismics is

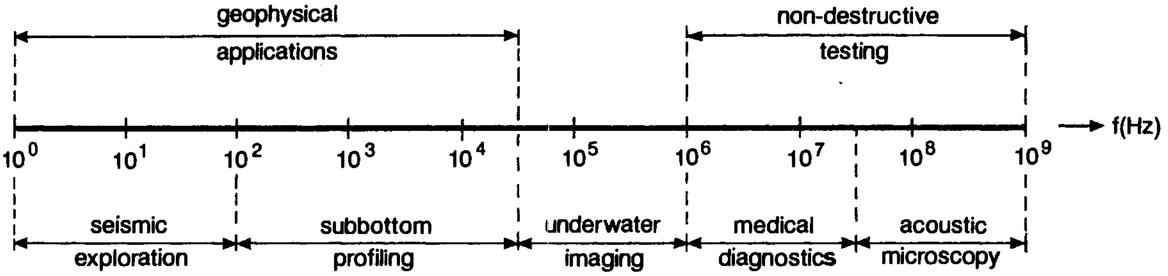


Fig. 4 Approximate frequency ranges as used in different acoustic imaging applications. Resolution increases with frequency; penetration decreases with frequency.

particularly used for hydrocarbon finding; the penetration is on the order of kilometres. High-frequency inspection occurs with the acoustic microscope; the penetration is on the order of millimetres.

In the following sections, we will concentrate on *seismic exploration*, that is, acoustic imaging of the geologic layers below the earth's surface. First, the principles will be reviewed in terms of ray theory (Section 3). Next, modern concepts are discussed in terms of wave theory (Section 4). Finally, a summary is given on seismic processing (Section 5).

3 PRINCIPLES OF SEISMIC EXPLORATION

Exploration seismology is based on analysis of acoustic waves reflected from different rock layers in the earth's subsurface. Impulsive acoustic energy, emitted into the subsurface, encounters discontinuities between the rock layers and is partially reflected back to the surface. The returning reflections are detected and their strengths and arrival times are recorded. After the recorded data are processed, geophysicists and geologists derive information from the result concerning the position, thickness, and rock property of each layer.

The seismic medium consists of layers of porous rock, such as sand, shale, and limestone. The pores of the rock generally contain water but may also contain gas and/or oil. In the oil and gas industry it is the objective of seismic exploration to identify structures that are favorable for containing a commercial amount of hydrocarbons.

Seismic reflection energy consists of compressional waves (P-waves) and shear waves (S-waves). Typical P-wave velocities are 500 m/s for the weathered surface layer at land (*land seismics*), 1500 m/s for the water surface layer at sea (*marine seismics*), 2500 m/s for consolidated sands and shales, and 4000 m/s for limestone layers. For most rock layers the ratio between P-wave and S-wave velocity (V_p–V_s ratio) varies between 1.5 and 2. This means that the Poisson ratio varies between 0.1 and 0.3, respectively. For shallow gas sands the V_p–V_s ratio may be as low as 1.4. For the weathered layer the V_p–V_s ratio may be as high as 2.5 (dry) to 5 (water filled). Of course, in the case of marine seismics $V_s = 0$ for the surface layer.

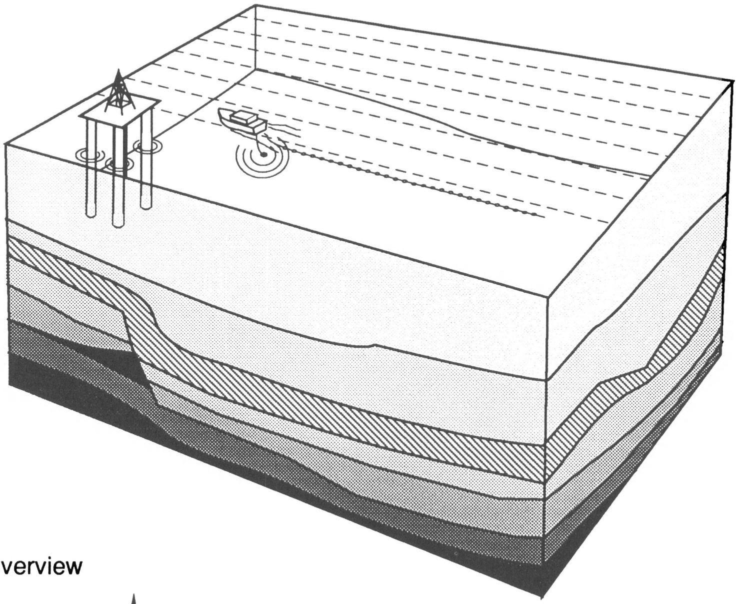
Due to the low-frequency content of seismic waves, the rock material and pore fluid move in phase as in a mixture. This means that the wave theory for *solids* may be used, the compressibility being given by the Gassmann equation.[2] The latter is consistent with the low-frequency approximation of the Biot theory.[3]

3.1 Seismic Acquisition

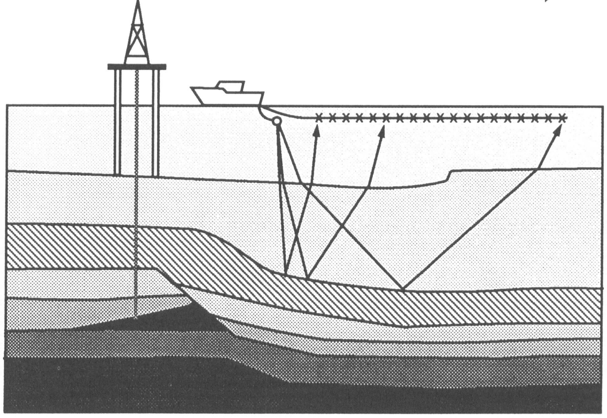
Figure 5 shows the conventional data acquisition geometry. The seismic source may consist of dynamite (land), a hydraulic vibrator (land and marine), or compressed air (marine). The subsurface response is measured by a number of detector stations. Each detector *station* consists of a number of electrically connected single detectors (*field pattern*). Each detector measures pressure (marine data) or the vertical component of the particle velocity (land data). The reflection sequence measured by one detector station is called a seismic trace. All traces related to one source position define one seismic *shot record*. Typically, there are 500 traces in one shot record. The data of one seismic line consist of a number of seismic shot records (at least several hundreds). Typical numerical quantities are as follows:

Distance between two adjacent shot positions	50 m
Distance between two adjacent detector–station positions	25 m
Length of a seismic trace	8 s
Time-sampling interval	4 ms

In a two-dimensional seismic survey the distance between adjacent seismic lines may be several kilometres. In a three-dimensional seismic survey adjacent seismic lines may be as close as 50 m. There is a general ten-



a) overview



b) cross section

Fig. 5 Conventional seismic data acquisition geometry; the marine single-streamer situation.

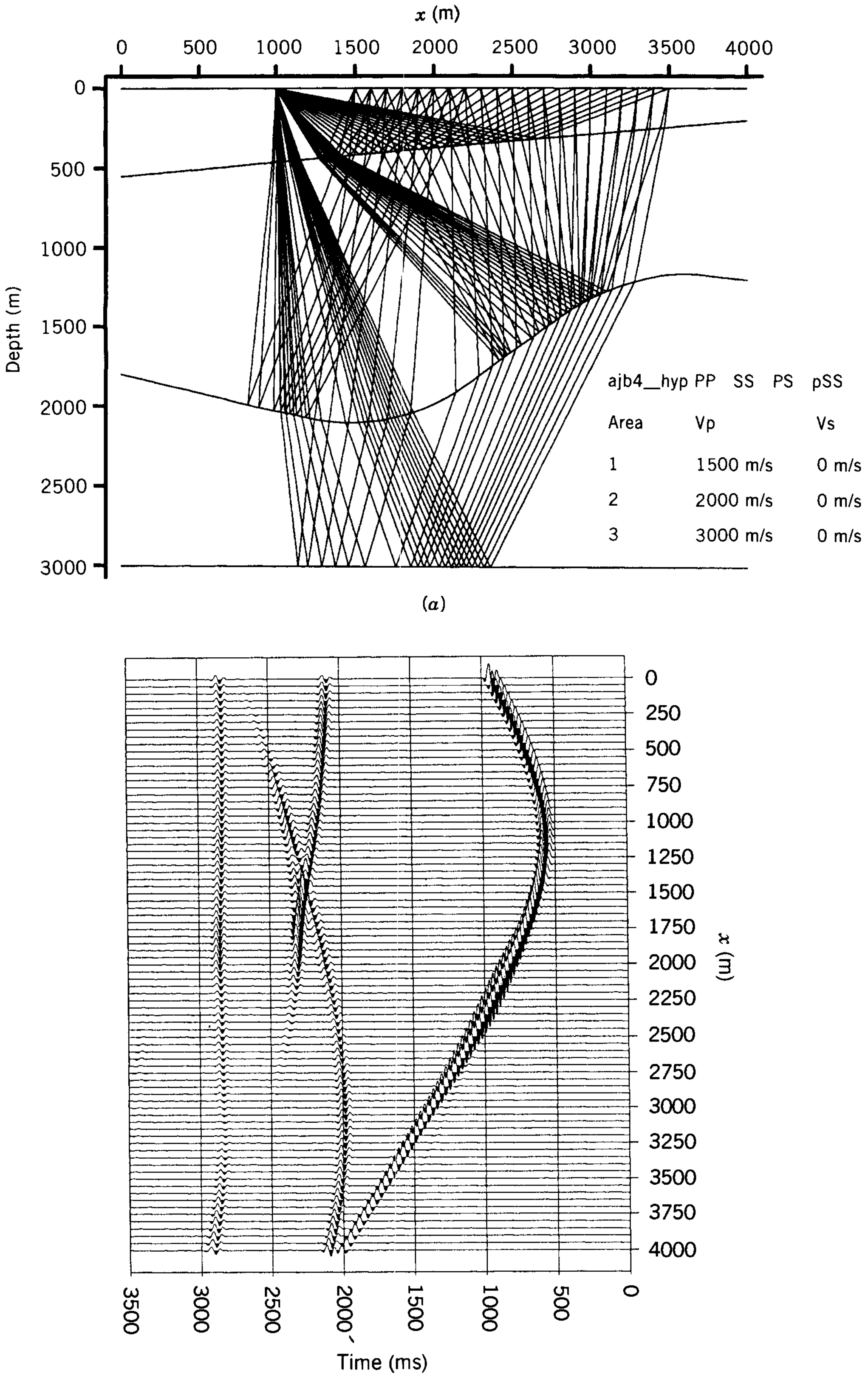
dency in the seismic industry to decrease shot and detector spacings, to decrease the length of field patterns, and to increase the number of detector stations.

Note that for a shot record of 500 traces and a recording length of 8 s the number of data bytes equals $500 \times 2000 \times 4 = 4.10^6$ (4 megabytes per shot record). Hence, for a three-dimensional seismic survey of 250,000 shots and total number of recorded bytes equals 10^{12} (1 terabyte per survey).

3.2 Seismic Processing

For the two-dimensional situation Fig. 6*a* illustrates with the aid of rays how a primary seismic wave field propagates through the subsurface (the multiple-scattering rays are not shown). The measurements at the surface define the shot record for the related source position (Fig. 6*b*). In the field the data are, after analog-to-digital conversion, stored on tape. Next, in the seismic processing centre data processing starts with several multioffset preprocessing steps: spatial resampling (*regularization*), removal of waves that travel *along* the surface (particularly for land data), attenuation of multiple scattering (particularly for marine data), correction for these source and receiver properties, and in case of land data, correction for the laterally irregular weathered surface layer.

Until now, the nucleus of any conventional seismic processing package consists of common midpoint (CMP) processing.[4] This means that *after* the above preprocessing steps the data are reordered such that all traces with the same midpoint between source and detector position are grouped in one gather, the so-called CMP gather. Figure 7*b* shows traces of one CMP gather. From the rays of the primary wave field (Fig. 7*a*) it can be seen that



(*a*)

(*b*)

Fig. 6 Seismic ray paths and seismic measurements related to one seismic experiment (shot record). For simplicity only primary longitudinal waves are shown.

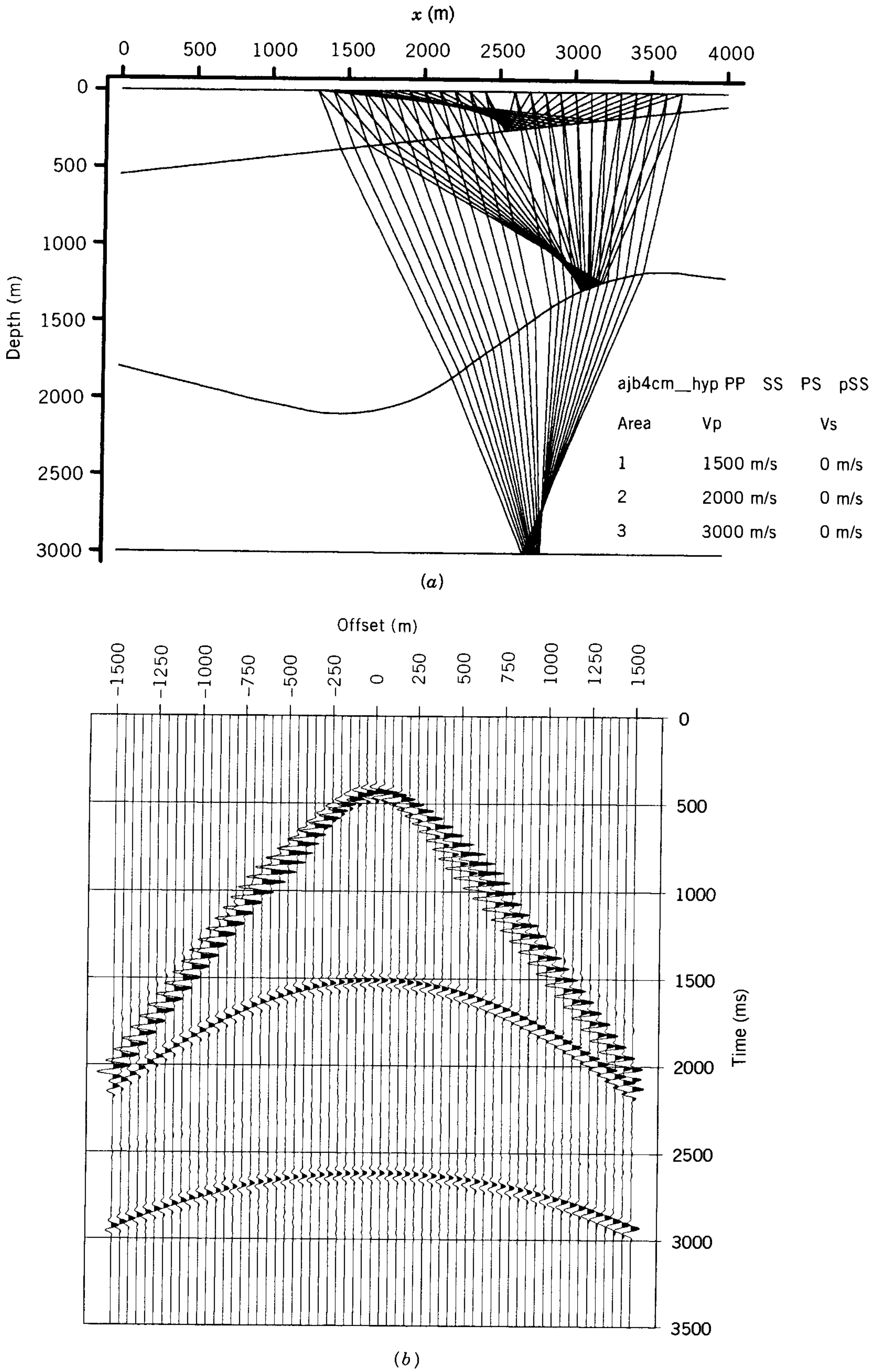


Fig. 7 Seismic ray paths and seismic measurements related to one CMP gather with its midpoint at $x = 2500$ m.

the reflection points for all traces in one CMP gather are grouped closely together. This property determines the usefulness of CMP processing. Even in fairly complicated geologic situations the arrival times of the reflections from one reflector in one CMP gather can be well approximated by a hyperbolic relationship,

$$t_{k,n}^2 = t_k^2 + \frac{(2n\,\Delta h)^2}{v_{\mathrm{st}}^2(t_k)}, \qquad (1)$$

where t_k is the zero-offset ($n = 0$) reflection time of the kth reflector and $h_n = 2n\,\Delta h$ represents the distance (*offset*) between source and detector of the nth trace in the CMP gather under consideration; $v_{\mathrm{st}}(t_k)$ is the so-called stacking velocity for the kth reflector.

Using a multitrace coherency criterion, v_{st} is estimated for a number of reflecting boundaries. From these values interval velocities for the layers between these boundaries can be computed. Figure 8 gives a contour plot of the coherency values related to the CMP data of Fig. 7*b*. Maxima are related to principal reflections. Together, they define v_{st} as a function of reflection time for the midpoint under consideration. For a comprehensive summary on the determination of stacking velocities the reader is referred to Ref. 5.

If one aims at high resolution in the $v_{\mathrm{st}}(t)$ space, use of the generalized radon transform is very effective. For more information the reader is referred to Ref. 6.

Given a stacking velocity distribution for a given CMP gather, the arrival times of all traces within this CMP gather are corrected (Fig. 9),

$$t_{k,n} \rightarrow t_k,$$

such that the travel time effect of the offset (*normal move-out*) is eliminated (*NMO correction*). Finally, all NMO-corrected traces are added together to form one so-called stack trace. In this way a considerable improvement in signal-to-noise ratio occurs and, last but not least, a significant data reduction is achieved. The *CMP stacking* process is characteristic of conventional seismic processing.

A stack trace is considered to be a zero-offset trace and all stack traces of a seismic line define a stack or pseudo-zero-offset section. Figure 10 shows a stack section together with the related zero-offset rays.

The pseudo-zero-offset section must be transformed to a zero-offset reflectivity image (Fig. 11). This means that each reflection is "migrated" to its own reflection point as defined by the endpoint of its zero-offset ray path. In addition, the reflection amplitude is corrected according to the propagation losses between surface and reflection point. The imaging process of repositioning and amplitude correction is called seismic migration, in this particular situation *zero-offset seismic migration*. To generate the migration operator for each reflection point, the macrovelocity model must be determined first. This can be done with the aid of the stacking velocity information.[7]

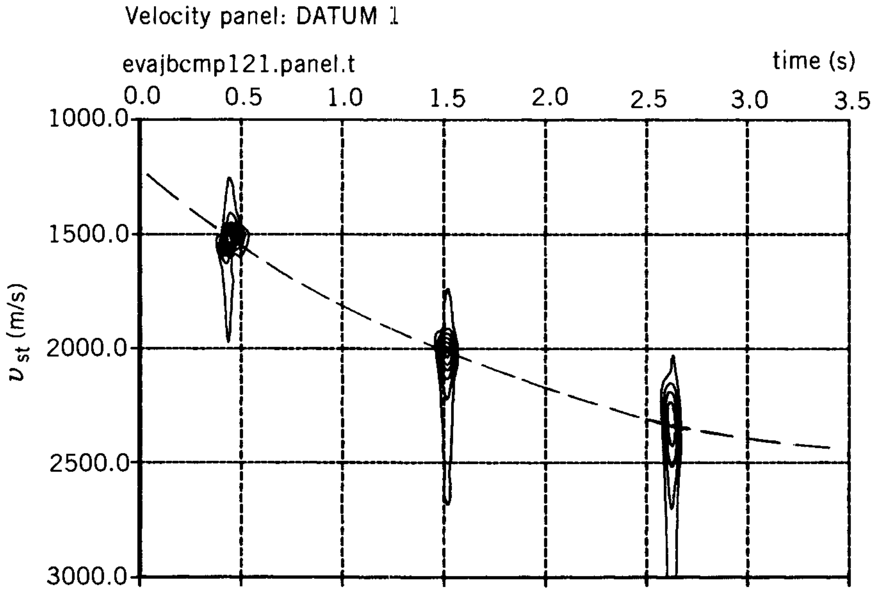


Fig. 8 Contour plot of the coherency values related to the CMP gather of Fig. 7 (velocity plot).

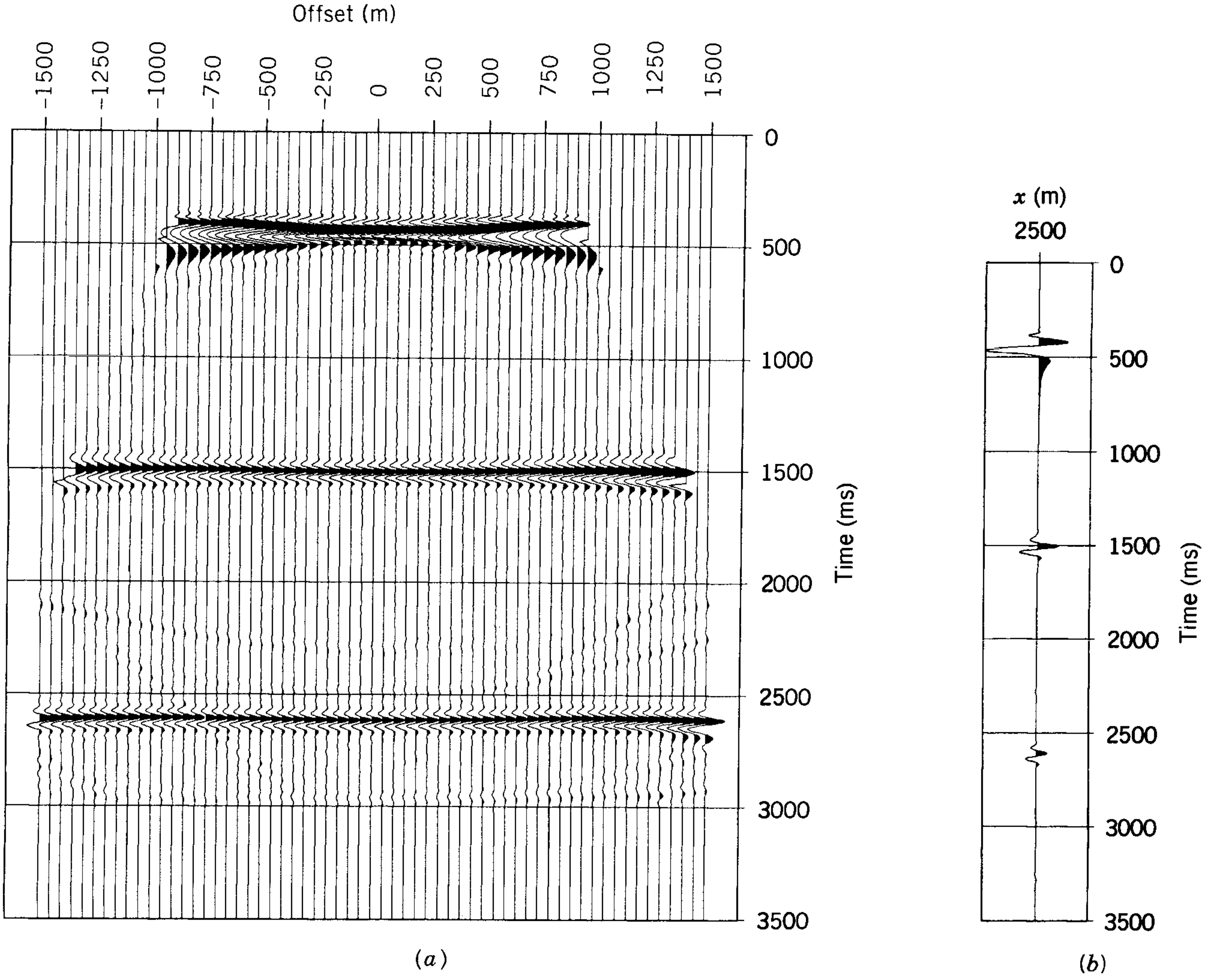


Fig. 9 The CMP gather of Fig. 7*b* after NMO (normal move-out) correction, together with its stacked version.

3.3 Seismic Interpretation

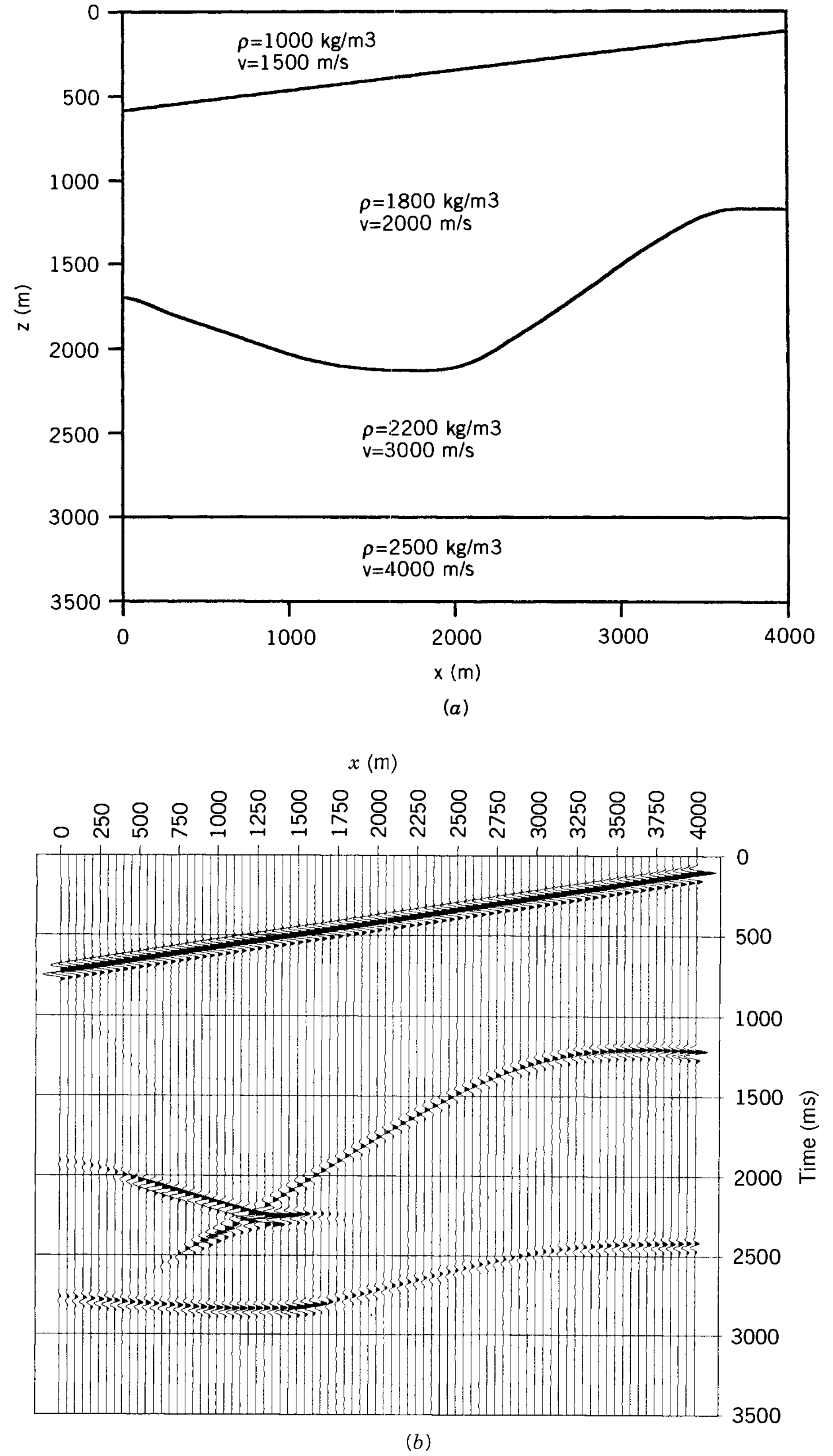
In the interpretation procedure the migrated reflections are tracked and integrated to a subsurface model. In this process the knowledge of *geological* processes (tectonics, sedimentation) is essential. In the early days tracking was pure hand work but nowadays the tracking process is largely automatic and carried out on interactive seismic workstations.

From the interpretation result potential hydrocarbon traps are indicated and well locations are proposed to verify and refine the seismic interpretation.

3.4 Summary

In the foregoing, the principles of seismic exploration have been explained in terms of ray theory. We have seen that, after data acquisition, the basic seismic processes multioffset preprocessing, CMP stacking, and zero-offset migration are applied (conventional method). In the preprocessing phase the attenuation of multiple scattering is an important step, particularly for marine data, where multiple scattering may be significant due to a large amount of trapped energy in the water layer. Actually, the CMP gather of Fig. 8 was already treated for multiple scattering. Figure 12*a* shows the same CMP gather, but now all multiples are included. Figure 12*b* illustrates the noticeable difference in stacking velocity between the primary reflection events and the multiples. This property can be used to remove the multiples.

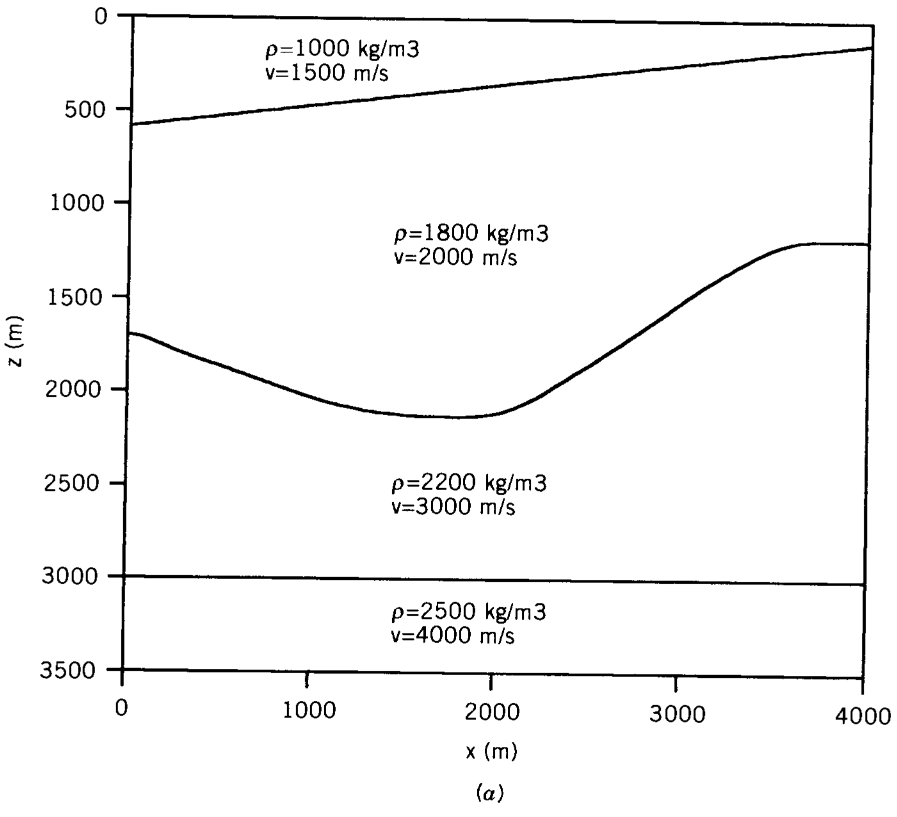
CMP stacking can be formulated as a numerical focusing process in the midpoint-offset domain, the result being positioned at the zero-offset times of each involved midpoint. Zero-offset migration can be formulated as a numerical focusing process in the midpoint

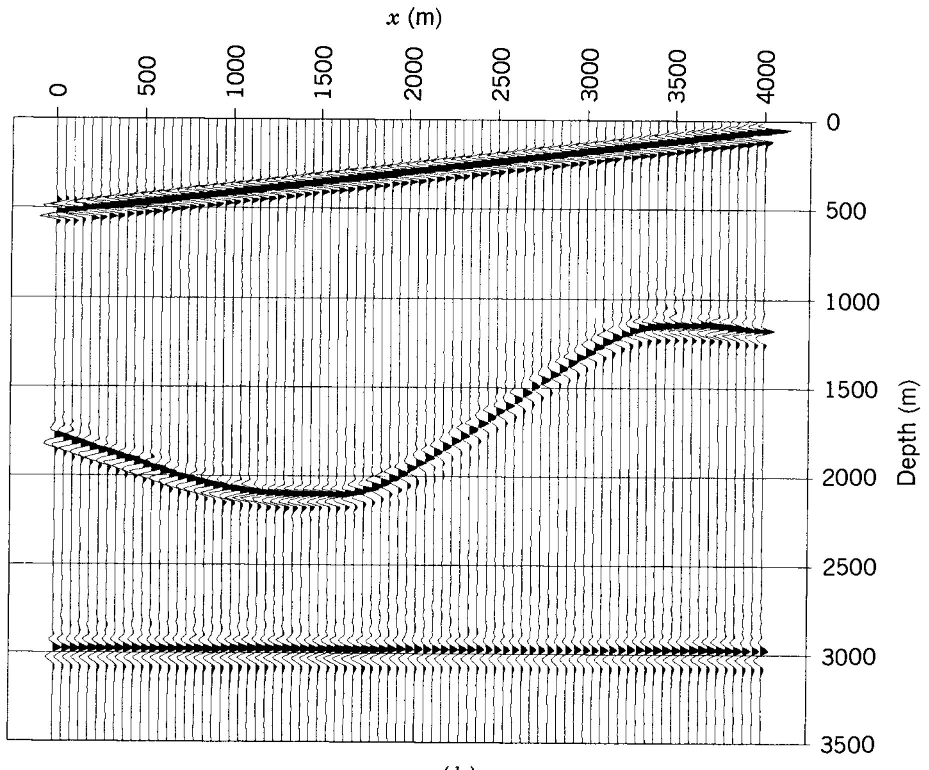


(*a*)

(*b*)

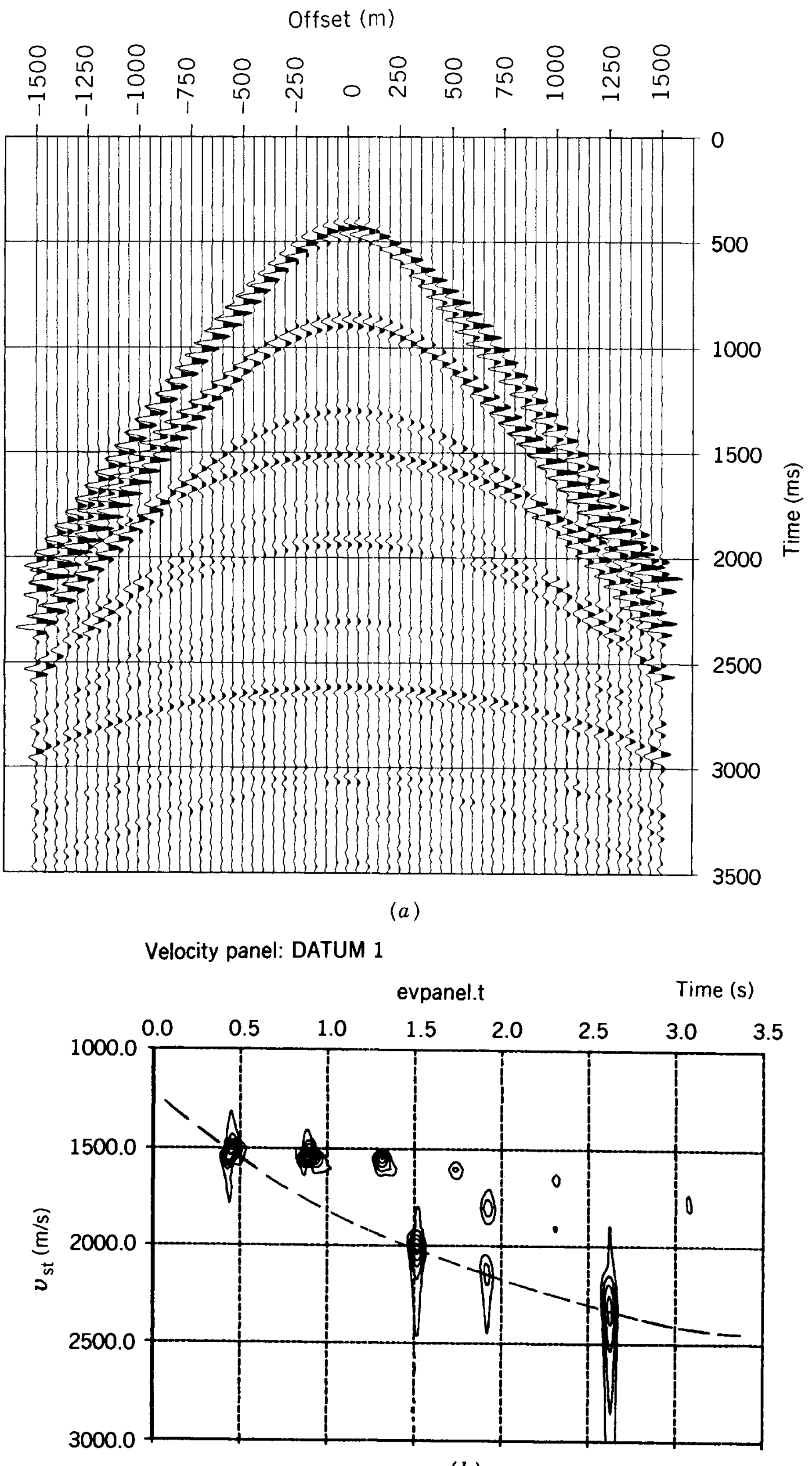
Fig. 10 Stack section (pseudo zero-offset section), consisting of all stacked CMP gathers: (*a*) subsurface model; (*b*) result of stacking. Note that CMP stacking defines a numerical focusing process.


ρ=1000 kg/m3
v=1500 m/s
ρ=1800 kg/m3
v=2000 m/s
ρ=2200 kg/m3
v=3000 m/s
ρ=2500 kg/m3
v=4000 m/s
z (m)
0
500
1000
1500
2000
2500
3000
3500
0
1000
2000
3000
4000
x (m)
(a)



(b)

Fig. 11 Migrated version of Fig. 10, yielding the zero-offset reflectivity image: (*a*) subsurface model; (*b*) result of migration. Note that zero-offset migration defines a numerical focusing process.



(*a*)

(*b*)

Fig. 12 Influence of multiple scattering on the CMP gather of Fig. 7: (*a*) CMP gather; (*b*) velocity analysis.

domain. Hence, conventional seismic processing consists of two consecutive focusing steps. The first focusing step translates multioffset measurements to zero-offset data (CMP stacking). The second focusing step translates zero-offset data to a subsurface image (zero-offset migration).

4 SEISMIC FORWARD MODEL

Today, the seismic exploration problem is approached from the *elastic wave theory* point of view: measurements are not considered as signals but as a superposition of waves. During preprocessing the different type of wave fields (surface wave field, multiple wave field, primary wave field) are separated, also referred to as wave field decomposition. Next, the pulse shape (*seismic wavelet*) of the primary wave field is shortened by model-based deconvolution and the output may be directly used in the image formation process (*multioffset migration*), yielding the structural information of the subsurface. In the target area the *angle-dependent* reflection property is estimated and used in the image characterization process (lithologic inversion), yielding the rock and pore information of the target.

In the following the seismic forward model is introduced in terms of a system of wave-theory-based operators.[8] We will see that the introduction of such a system makes the discussion on the seismic processing problem very straightforward.

4.1 Measurement Matrix

In practice acoustical measurements are always discrete in time and space. Consequently, imaging is always a discrete process and the theory should be discrete. Therefore, the forward model for seismic measurements is presented as a *discrete* model.

For linear wave theory in a time-invariant medium the problem may be described in the temporal frequency domain. Moreover, as our recording has a finite duration (T), we are dealing with a finite number of frequencies (N) per recorded signal (seismic trace), where

$$N = (f_{\max} - f_{\min})T,$$

$f_{\max} - f_{\min}$ being the temporal frequency range of interest. A typical number for N is 250. Taking into account the discrete property on the one hand and the allowed representation by independent frequency components on the other hand, vectors and matrices are preeminently suited for the mathematical description of seismic measurements. For instance, considering one seismic experiment (one source position), one element of the so-called measurement vector $\mathbf{P}(z_0)$ contains the complex number (defining amplitude and phase for the Fourier component under consideration) related to the recorded signal at one location of acquisition plane $z = z_0$ (one detector position). For a number of experiments the related measurement vectors can be grouped in a so-called measurement *matrix* (Fig. 13). Note that the main diagonal represents zero-offset data, the off-diagonals represent common offset data and the antidiagonals represent common midpoint data. In the next section the forward model will be derived to explain the data in the seismic measurement matrix. The symbol z will be used for depth.

4.2 Forward Model for Reflection Measurements

If the vector $\mathbf{S}^+(z_0)$ represents one Fourier component of the downward-traveling source wave field at the data acquisition surface $z = z_0$, then we may write

$$\mathbf{P}^+(z_m) = \boldsymbol{W}^+(z_m, z_0)\mathbf{S}^+(z_0), \tag{2}$$

where the vector $\mathbf{P}^+(z_m)$ represents the monochromatic downward-traveling source wave field at depth level z_m and $\boldsymbol{W}^+(z_m, z_0)$ defines the downward propagation operator from z_0 to z_m. Operator $\boldsymbol{W}^+$ is represented by a complex-valued matrix, where each column equals one Fourier component of the response at depth level z_m due to one dipole at the surface. Note that for one-dimensional media (no lateral variations) $\boldsymbol{W}^+$ becomes a convolution matrix.

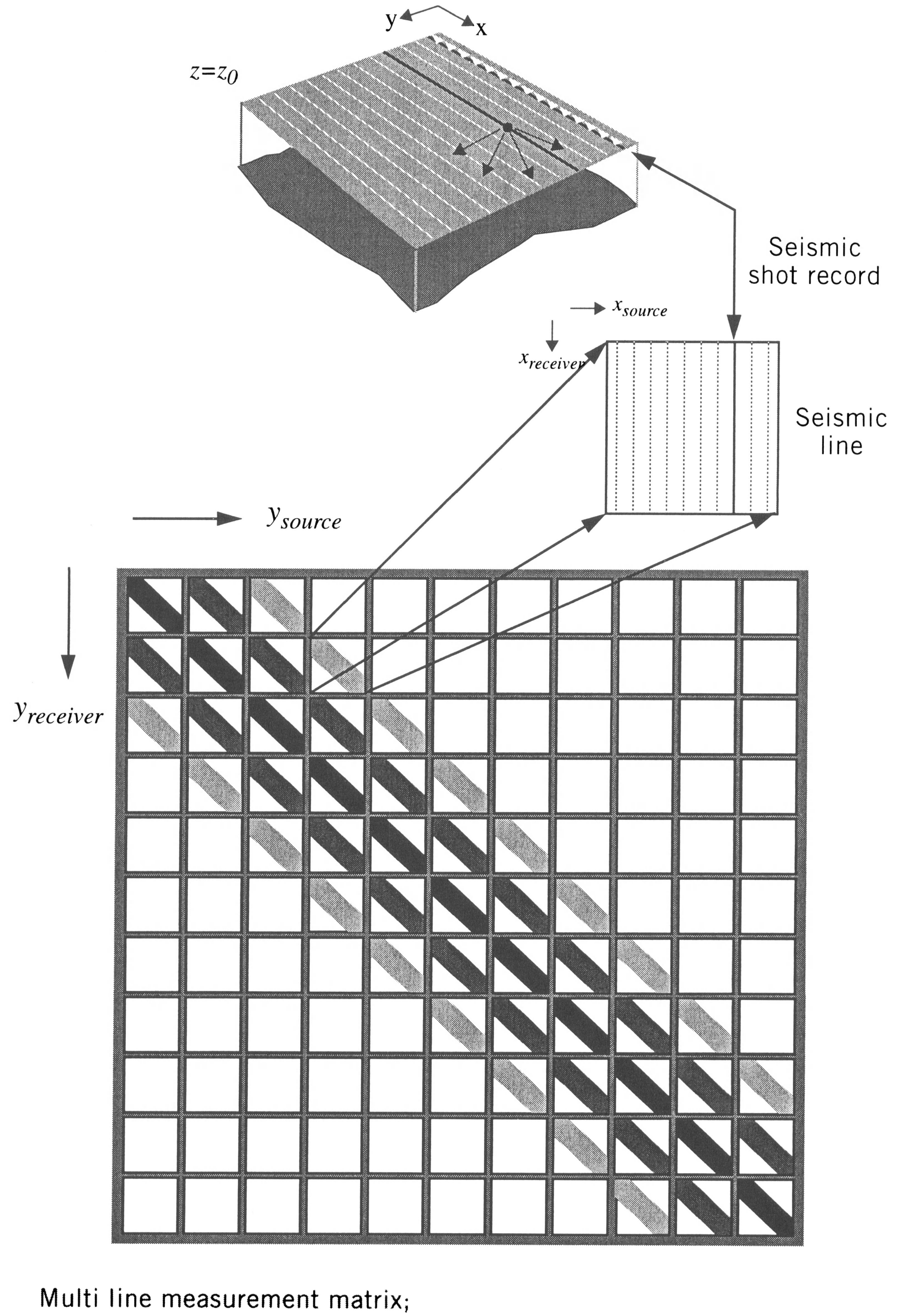
At depth level z_m reflection occurs. For each Fourier component reflection at z_m may be described by a general linear operator $\boldsymbol{R}^+(z_m)$,

$$\mathbf{P}_m^-(z_m) = \boldsymbol{R}^+(z_m)\mathbf{P}^+(z_m), \tag{3}$$

where the vector $\mathbf{P}_m^-(z_m)$ represents the monochromatic upward-traveling reflected wave field at depth level z_m due to the inhomogeneities at depth level z_m only. Reflection operator $\boldsymbol{R}^+(z_m)$ is represented by a matrix, where each column describes the monochromatic angle-dependent reflection property of each grid point at z_m. If there exists no angle dependence, $\boldsymbol{R}^+(z_m)$ is a diagonal matrix with angle-independent reflection coefficients. Finally, the reflected wave field at z_m travels up to the surface,

$$\mathbf{P}_m^-(z_0) = \boldsymbol{W}^-(z_0, z_m)\mathbf{P}_m^-(z_m). \tag{4}$$

In Eq. (4) vector $\mathbf{P}_m^-(z_0)$ represents one Fourier component of the reflected wave field at data acquisition surface z_0 and $\boldsymbol{W}^-(z_0, z_m)$ defines the upward propagation operator from z_m to z_0. Each column of the matrix $\boldsymbol{W}^-$ contains


y
x
z=z0
Seismic shot record
x source
x receiver
Seismic line
y source
y receiver

Multi line measurement matrix;
for $Y_{source} \neq Y_{receiver}$ the source has an offset with respect to the receiver line.

Fig. 13 Data matrix for three-dimensional seismic measurements. One column represents one multistreamer shot record. For an irregular dataset the columns are irregularly filled.

one Fourier component of the response at z_0 due to one dipole at depth level z_m.

Expressions (2), (3), and (4) may now be combined to one matrix equation for the reflection response:

$$
\begin{aligned}
\mathbf{P}^-(z_0) &= \sum_{m=1}^{\infty} \mathbf{P}^-_m(z_0) \\
&= \sum_{m=1}^{\infty} \boldsymbol{W}^-(z_0, z_m)\mathbf{P}^-_m(z_0) \\
&= \sum_{m=1}^{\infty} \boldsymbol{W}^-(z_0, z_m)\boldsymbol{R}^+(z_m)\mathbf{P}^+(z_m) \\
&= \sum_{m=1}^{\infty} [\boldsymbol{W}^-(z_0, z_m)\boldsymbol{R}^+(z_m) \\
&\quad \cdot \boldsymbol{W}^+(z_m, z_0)]\mathbf{S}^+(z_0), \qquad (5a)
\end{aligned}
$$

or, for a continuous formulation in z,

$$\mathbf{P}^-(z_0) = \int_{z_0}^{\infty} [\boldsymbol{W}^-(z_0, z)\boldsymbol{R}^+(z)\boldsymbol{W}^+(z, z_0)dz]\mathbf{S}^+(z_0). \qquad (5b)$$

Note that in expression (5) the surface reflectivity is not yet included [$\mathbf{R}^+(z_0) = 0$]. Note also that Eq. (5) contains the primary reflection response only; multiples and surface waves have been omitted.

If we define the half-space reflection operator at depth level z_m due to inhomogeneities at $z \geq z_m$ by matrix $\boldsymbol{X}^+(z_m, z_m)$, then it follows from Eq. (5a) that

$$\boldsymbol{X}^+(z_m, z_m) = \sum_{n=m}^{\infty} \boldsymbol{W}^-(z_m, z_n)\boldsymbol{R}^+(z_n)\boldsymbol{W}^+(z_n, z_m), \qquad (6a)$$

where $m = 1, 2, \ldots$.

Matrix $\boldsymbol{X}^+(z_m, z_m)$ may be considered as a multidimensional *transfer function*; matrix element $\mathbf{X}^+_{ij}(z_m, z_m)$ represents one Fourier component of the reflection response at position i on surface $z = z_m$, due to a unit point source at position j on the same surface ($z = z_m$).

Finally, if we define

$$\boldsymbol{X}^+_0(z_m, z_m) = \sum_{n=m+1}^{\infty} \boldsymbol{W}^-(z_m, z_n)\boldsymbol{R}^+(z_n)\boldsymbol{W}^+(z_n, z_m), \qquad (6b)$$

then it follows from Eq. (5) that

$$\mathbf{P}^-(z_0) = \boldsymbol{X}^+_0(z_0, z_0)\mathbf{S}^+(z_0). \qquad (6c)$$

4.3 Including the Acquisition Surface

To expressions (5) the relations should be added between the *induced-source function* and the downgoing wave field on the one hand and the recorded detector signals and the upgoing wave field on the other hand:

$$
\begin{aligned}
\mathbf{S}^+(z_0) &= \boldsymbol{D}^+(z_0)\mathbf{S}(z_0), \qquad (7a) \\
\mathbf{P}(z_0) &= \boldsymbol{D}^-(z_0)\mathbf{P}^-(z_0). \qquad (7b)
\end{aligned}
$$

The matrix operators $\boldsymbol{D}^+$ and $\boldsymbol{D}^-$ represent the local interaction with the surface, the directivity properties due to the array effects, and the type of sources and detectors, respectively.

In the seismic situation the data acquisition surface is the earth's surface. It is highly reflective and its significant influence on the seismic response can be easily included in the forward model by replacing in expression (5) downgoing source wave field $\mathbf{S}^+(z_0)$ by

$$\mathbf{P}^+(z_0) = \boldsymbol{R}^-(z_0)\mathbf{P}^-(z_0) + \mathbf{S}^+(z_0), \qquad (7c)$$

where the vector $\mathbf{P}^+(z_0)$ represents the *total downgoing* wave field at the surface and matrix $\boldsymbol{R}^-(z_0)$ defines the reflectivity of the surface for *upgoing* waves (Fig. 14). Expression (7c) formulates a special (and for practical seismic situations the most important) form of multiple scattering.

Finally, if we include noise in the measurements, then Eq. (7b) needs to be generalized to

$$\mathbf{P}(z_0) = \boldsymbol{D}^-(z_0)\mathbf{P}^-(z_0) + \mathbf{N}(z_0). \qquad (7d)$$

Here, noise is defined as all events that are not described by the model equations (e.g., waves that travel from source to detectors *along* the surface).

In Fig. 14 the foregoing theory is summarized in a diagram, showing the involved operators and data flow. Not

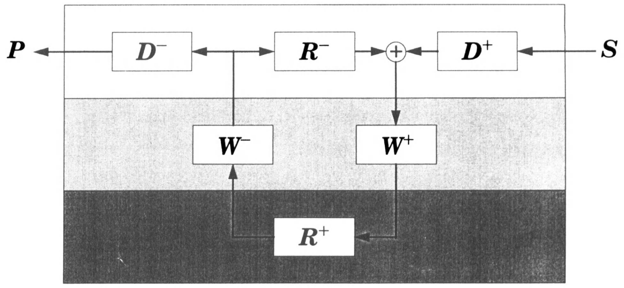


Fig. 14 Conceptual model of seismic data, consisting of surface operators ($\boldsymbol{D}^+$, $\boldsymbol{R}^-$, $\boldsymbol{D}^-$), propagation operators ($\boldsymbol{W}^+$, $\boldsymbol{W}^-$), and reflectivity operator $\boldsymbol{R}^+$.

only does the presented *conceptual* view on the problem yield a concise formulation but also the involved expressions provides theoretical detail. In addition, the involved basic algorithms are included if the operators are represented by matrices with predefined structure (Fig. 13). For instance, multiplication of a seismic dataset with the matrix operator $\boldsymbol{W}^+$ (or $\boldsymbol{W}^-$) defines a *forward* wave field extrapolation process, the (optimized) extrapolation operator for lateral position x_k being represented by the kth column of $\boldsymbol{W}^+$ (or $\boldsymbol{W}^-$). Similarly, multiplication of a seismic dataset with matrix operator $\boldsymbol{R}^+$ defines an angle-dependent reflection process, the (optimized) reflection operator for lateral position x_k being represented by the kth column of $\boldsymbol{R}^+$. In the opinion of the author, conceptual formulations are the way that experts of different disciplines should communicate with each other in interdisciplinary project teams. Moreover, as conceptual formulations are principally different from simplified ones, conceptual formulations appear be very fruitful for the development of new algorithms.[8] Examples of particular interest are surface-related multiple removal[9] and controlled illumination.[10]

4.4 Multiexperiment Data Sets

So far, we have formulated the equations for one experiment. However, our matrix formulation is very suited to represent multiexperiment data acquisition:

$$\boldsymbol{S}^+(z_0) = \boldsymbol{D}^+(z_0)\boldsymbol{S}(z_0), \tag{8a}$$

$$\boldsymbol{P}^+(z_0) = \boldsymbol{R}^-(z_0)\boldsymbol{P}^-(z_0) + \boldsymbol{S}^+(z_0), \tag{8b}$$

$$\boldsymbol{P}^-(z_0) = \sum_{m=1}^{\infty} [\boldsymbol{W}^-(z_0, z_m)\boldsymbol{R}^+(z_m) \cdot \boldsymbol{W}^+(z_m, z_0)]\boldsymbol{P}^+(z_0), \tag{8c}$$

$$\boldsymbol{P}(z_0) = \boldsymbol{D}^-(z_0)\boldsymbol{P}^-(z_0) + \boldsymbol{N}(z_0), \tag{8d}$$

where one column of the source matrix $\boldsymbol{S}(z_0)$ defines the induced source function of one monochromatic experiment and the related column of the measurement matrix $\boldsymbol{P}(z_0)$ defines the monochromatic versions of the measured signals (seismic traces) of that experiment.

If the source directivity and the surface-related multiple scattering have been removed, then we may write $\boldsymbol{P}^+(z_0) = S(\omega)\boldsymbol{I}$, and it can be easily verified from Eq. (8c) that the elements of $\boldsymbol{P}^-(z_0)$ can be written as

$$P^-_{ij}(z_0) = S(\omega)\sum_m\sum_k\sum_l \cdot W^-_{ik}(z_0, z_m)R^+_{kl}(z_m)W^+_{lj}(z_m, z_0), \tag{9a}$$

$P^-_{ij}(z_0)$ being the reflected wave field at surface position i due to a dipole at surface position j. Hence for zero-offset data ($i = j$), where each grid point may be characterized by a zero-offset reflection coefficient ($k = l$), we may write

$$P^-_{ii}(z_0) = S(\omega)\sum_{m=1}^{\infty}\sum_k W_{ik}(z_0, z_m)R^+_{kk}(z_m). \tag{9b}$$

In Eq. (9b) we have defined

$$W_{ik}(z_0, z_m) = W^-_{ik}(z_0, z_m)W^+_{ki}(z_m, z_0) \tag{9c}$$

and $R^+_{kk}(z_m)$ equals the zero-offset reflection coefficient at grid point x_k of depth level z_m.

4.5 Formulation for both P- and S-Waves

As mentioned before, seismic wave fields contain P as well as S reflections and, therefore, the seismic forward model should include both type of waves. This is particularly important for land data. We will see that the forward model (8) can be elegantly extended to address both P and S reflections:

(a) Each scalar element of $\boldsymbol{S}(z_0)$ should be replaced by a vector element representing the source stress components τ_{xz}, τ_{yz}, and τ_{zz}.

(b) Each scalar element of $\boldsymbol{P}(z_0)$ should be replaced by a vector element representing the detector particle velocity components v_x, v_y, v_z.

(c) Each scalar element of $\boldsymbol{S}^+(z_0)$ should be replaced by a vector element representing the P and S potentials ϕ^+, Ψ^+_x, Ψ^+_y at z_0. This means that surface operator $\boldsymbol{D}^-(z_0)$ decomposes the downgoing source wave field in separate P and S contributions as well.

(d) Each scalar element of $\boldsymbol{P}^-(z_0)$ should be replaced by a vector element, representing the P and S potentials ϕ^-, Ψ^-_x, Ψ^-_y at z_0. This means that surface operator $\boldsymbol{D}^-(z_0)$ composes the upgoing P and S wave fields in particle velocity components as well.

Taking the above generalizations of Eqs. (8a) and (8d) into account, expressions (8b) and (8c) can be easily generalized as well. For simplicity the two-dimensional situation is formulated:

$$\overset{\substack{\mathbf{P}^+(z_0)\\ \downarrow}}{\begin{bmatrix}\phi^+\\ \Psi^+\end{bmatrix}} = \overset{\substack{\mathbf{R}^-(z_0)\\ \downarrow}}{\begin{bmatrix}\mathbf{R}^-_{pp} & \mathbf{R}^-_{ps}\\ \mathbf{R}^-_{sp} & \mathbf{R}^-_{ss}\end{bmatrix}} \overset{\substack{\mathbf{P}^-(z_0)\\ \downarrow}}{\begin{bmatrix}\phi^-\\ \Psi^-\end{bmatrix}} + \overset{\substack{\mathbf{S}^+(z_0)\\ \downarrow}}{\begin{bmatrix}S^+_\phi\\ S^+_\Psi\end{bmatrix}}, \tag{10a}$$

$$\underset{\substack{\uparrow\\ \mathbf{P}^-(z_0)}}{\begin{bmatrix}\phi^-\\ \Psi^-\end{bmatrix}} = \sum_{m=1}^{\infty} \underset{\substack{\uparrow\\ \mathbf{W}^-(z_0,z_m)}}{\begin{bmatrix}\mathbf{W}^-_{pp} & \mathbf{W}^-_{ps}\\ \mathbf{W}^-_{sp} & \mathbf{W}^-_{ss}\end{bmatrix}} \underset{\substack{\uparrow\\ \mathbf{R}^+(z_m)}}{\begin{bmatrix}\mathbf{R}^+_{pp} & \mathbf{R}^+_{ps}\\ \mathbf{R}^+_{sp} & \mathbf{R}^+_{ss}\end{bmatrix}}$$

$$\cdot \underset{\substack{\uparrow\\ \mathbf{W}^+(z_m,z_0)}}{\begin{bmatrix}\mathbf{W}^+_{pp} & \mathbf{W}^+_{ps}\\ \mathbf{W}^+_{sp} & \mathbf{W}^+_{ss}\end{bmatrix}} \underset{\substack{\uparrow\\ \mathbf{P}^+(z_0)}}{\begin{bmatrix}\phi^+\\ \Psi^+\end{bmatrix}}. \tag{10b}$$

Note that the matrices $\mathbf{R}_{sp}$ and $\mathbf{R}_{ps}$ quantify wave type conversion during *reflection*. In many cases it is acceptable to neglect wave type conversion during *propagation*, meaning that $\mathbf{W}_{sp}$ and $\mathbf{W}_{ps}$ may be neglected. Note also that in seismic marine exploration $\Psi = 0$ in the surface layer and receivers detect pressure just below z_0.

In conventional seismic land acquisition, stress sources generate τ_{zz} only (vertical vibrator) and detectors measure v_z only (single-component geophone). In the common situation of a low velocity surface layer ray paths are near to vertical at the detectors and, therefore, *vertical* component geophones measure mainly *compressional* reflection energy.

It can be easily verified that if $\mathbf{W}_{sp}$ and $\mathbf{W}_{ps}$ are neglected, Eq. (10b) can be written as four independent equations (two without conversion and two with conversion at reflection). This has the important practical advantage that the multicomponent *elastic* problem can be formulated in terms of four (two-dimensional) or nine (three-dimensional) independent single-component *acoustic* problems.

5 SEISMIC IMAGING

In the *forward* problem all details about the data acquisition procedure are known, the acoustic properties of the surface and medium (trend and detail) are available, and the multioffset measurements need be computed (numerical simulation). In case we start with reflectivity, simulation means

$$\mathbf{R}^+(z) \rightarrow \mathbf{X}_0^+(z_0, z_0) \rightarrow \mathbf{P}(z_0).$$

In the *inverse* problem, all details about the data acqusition procedure should be known, the measurements are available, and the medium parameters need to be computed. If the spatial reflectivity distribution is aimed for, inversion means multioffset preprocessing and multioffset image formation:

$$\mathbf{P}(z_0) \rightarrow \mathbf{X}_0^+(z_0, z_0) \rightarrow \mathbf{R}^+(z).$$

Generally, in image formation the *diagonal* elements of $\mathbf{R}^+(z)$ are computed only, meaning that the angle dependence information of reflection is not aimed for (one reflection coefficient per medium grid point).

If all elements of the reflection matrix are computed, then for each grid point angle-dependent reflection information is available and image *formation* can be followed by image *characterization*:

$$\mathbf{R}^+(z) \rightarrow \text{lithology}.$$

Lithologic information includes types of rock (sand, shale, etc.), types of pore fill (water, oil, gas), porosity, and permeability.

Hence, full seismic processing consists of the three hierarchical inversion steps, i.e., multioffset preprocessing, multioffset image formation, and multioffset image characterization:

$$\mathbf{P}(z_0) \rightarrow \mathbf{X}_0^+(z_0, z_0) \rightarrow \mathbf{R}^+(z) \rightarrow \text{lithology}.$$

5.1 Preprocessing

From the forward model (8) it can be seen that if both the total measurement matrix and the total source matrix are available (multisource, multioffset data acquisition), then the transfer matrix $\mathbf{X}_0^+(z_0, z_0)$ can be estimated by removing the following surface-related influences from the measurement matrix $\mathbf{P}(z_0)$:

(a) Surface wave field $\mathbf{N}(z_0)$
(b) Transducer operators $\mathbf{D}^-(z_0)$, $\mathbf{D}^+(z_0)$
(c) Induced source function $S(\omega)$
(d) Multiple scattering operator $\mathbf{R}^-(z_0)$.

Removal of the surface wave field (particularly for land measurements) occurs with the least-squares linear radon transformation. For details the reader is referred to Ref. 11.

Removal of transducer operators $\mathbf{D}^+(z_0)$ and $\mathbf{D}^-(z_0)$ transforms the source functions into *downgoing* wave fields at z_0 and the measurements into *upgoing* reflected wave fields at z_0. In addition, directivity effects are compensated for. For details see Ref. 12.

Removal of $S(\omega)$ is referred to as seismic source deconvolution; it means pulse compression, improving the seismic resolution along the time axis. Seismic source deconvolution is limited by the signal-to-noise ratio and the accuracy of the source signal estimate (see, Refs. 13 and 14).

Removal of $\boldsymbol{R}^-(z_0)$ means elimination of surface-related multiple scattering. From Eq. (8) it can be easily verified that the elimination process is defined by

$$\boldsymbol{P}_0(z_0) = \boldsymbol{P}(z_0)[\boldsymbol{I} + \boldsymbol{A}(z_0)\boldsymbol{P}(z_0)]^{-1}, \qquad (11a)$$

$\boldsymbol{P}_0(z_0)$ representing the seismic data without surface-related multiples and

$$\boldsymbol{A}(z_0) = [\boldsymbol{S}^+(z_0)]^{-1}\boldsymbol{R}^-(z_0)[\boldsymbol{D}^-(z_0)]^{-1}. \qquad (11b)$$

For details the reader is referred to Ref. 9

Figure 15 shows a result of preprocessing. The broadband version of one column of the matrix $\boldsymbol{P}(z_0)$ is shown before and after preprocessing (data from the North Sea). Note the significant amount of multiple scattering energy.

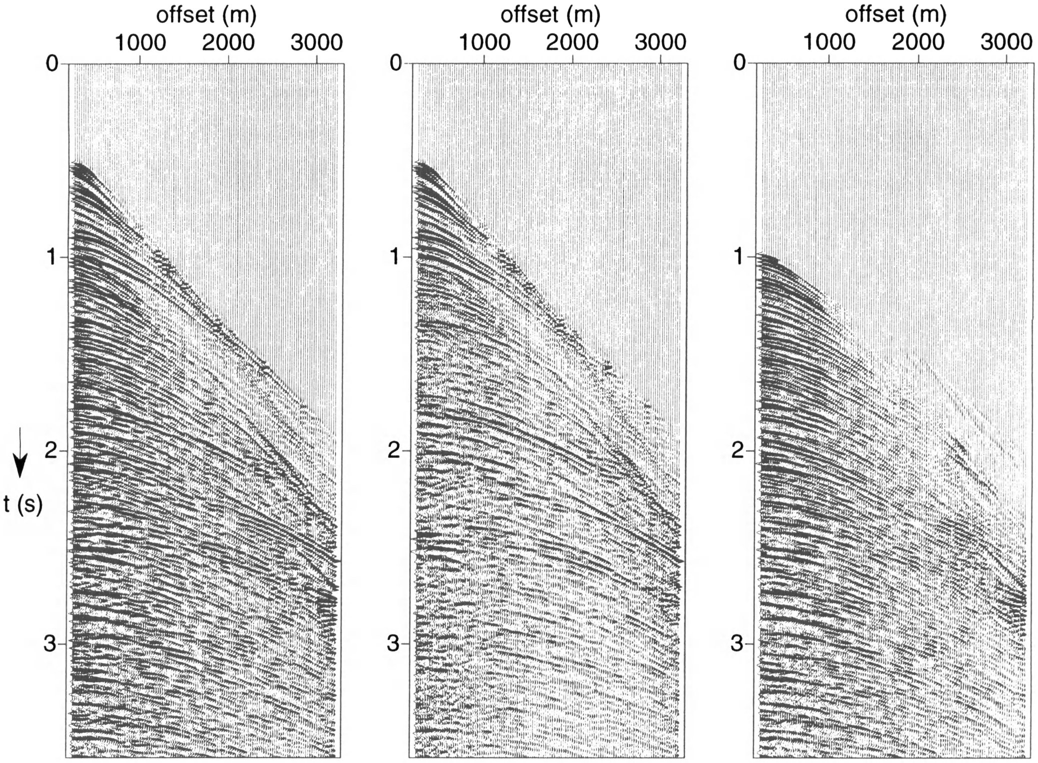
5.2 Image Formation

Removal of propagation operators $\boldsymbol{W}^+(z_m, z_0)$ and $\boldsymbol{W}^-(z_0, z_m)$ transforms $\boldsymbol{X}_0^+(z_0, z_0)$ to $\boldsymbol{X}^+(z_m, z_m)$; it means simulation of the data acquisition process at depth level z_m, $m = 1, 2, \ldots$. In the seismic literature this numerical process is referred to as *downward extrapolation*. From $\boldsymbol{X}^+(z_m, z_m)$ the reflection matrix $\boldsymbol{R}^+(z_m)$ can be easily extracted. If the diagonal elements of $\boldsymbol{R}^+(z_m)$ are computed only, then the twofold downward extrapolation process can be rewritten as double numerical focusing, simulating dynamic focusing in emission and dynamic focusing in detection for each grid point at depth level z_m. This is exactly the process that is carried out in multioffset migration. The multioffset migration process can be considered as the most advanced image formation process, addressing both the multisource and multioffset problems and avoiding the simplifying assumptions of CMP stacking in conventional seismic processing. For further reading see Refs. 8 and 15–17.

Figure 16 shows a result of the multioffset migration process using all columns of the preprocessed North Sea measurement matrix. Note the well-defined faults at a depth of 3 km.

5.3 Image Characterization

In the last inversion step of seismic processing angle-dependent reflection information, as given by $\boldsymbol{R}^+(z)$, is



a) input *b) output* *c) difference*

Fig. 15 One seismic experiment (*a*) before and (*b*) after removal of multiple scattering. (*c*) Note the significant amount of multiple scattering energy.

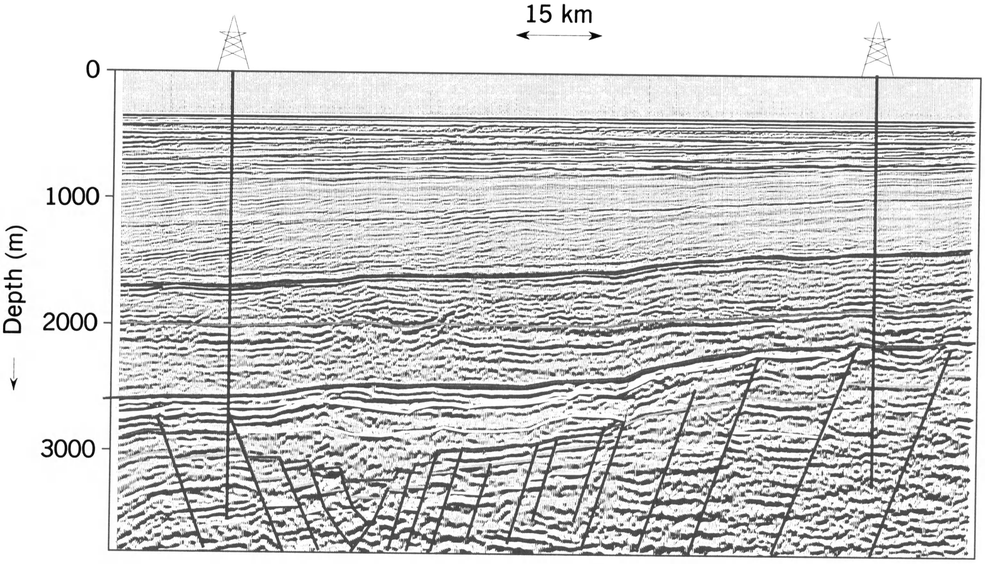


Fig. 16 Seismic image of the subsurface after preprocessing and multisource-multioffset image formation. Note that the latter may be formulated as a double focusing process: focusing in emission and focusing in detection.

inverted to a layered geological model, each layer being characterized by its rock and pore parameters. For this inversion step three forward models are used:

(a) Boundary conditions for angle-dependent reflection, e.g., as given by the Zoeppritz equations for solids[18]

(b) Gassmann equation for the P and S velocities of porous rocks[2]

(c) Density equation for porous rocks.

Using a constrained data-fitting technique,[19] the parameters of the forward models are adjusted such that the seismic reflectivity image(s) are explained as well as possible. Experience of seismic inversion for litho-parameters shows that the problem is seriously ill-posed. Success can only be realized if other, nonseismic, information is included as well (multidisciplinary approach).

Lithologic inversion needs, apart from seismic data, borehole information and a range of plausible geological models. The inversion process chooses the most likely one, together with limits of confidence.[20] Lithologic inversion is still largely in the research phase.

REFERENCES

1. A. J. Berkhout, *Seismic Resolution: Resolving Power of Acoustic Echo Techniques*, Geophysical Press, London, 1984.
2. F. Gassmann, "Elastic Waves through a Packing of Spheres," *Geophysics*, Vol. 16, 1951, pp. 673–685.
3. M. A. Biot, "Theory of Propagation of Elastic Waves in a Fluid Saturated Porous Solid," *J. Acoust. Soc. Am.*, Vol. 28, 1956, pp. 168–178.
4. O. Yilmaz, *Seismic Data Processing*, Society of Exploration Geophysicists, Tulsa, OK, 1987.
5. M. T. Taner and F. Koehler, "Velocity Spectra—Digital Computer Derivation and Application of Velocity Functions," *Geophysics*, Vol. 34, 1969, pp. 859–881.
6. D. Hampson, "Inverse Velocity Stacking for Multiple Elimination," *J. CSEG*, Vol. 22, No. 1, 1986, pp. 44–55.
7. P. Hubral, "Interval Velocities from Surface Measurements in the Three Dimensional Plane Layer Case," *Geophysics*, Vol. 41, 1976, pp. 223–242.
8. A. J. Berkhout, *Seismic Migration: Imaging of Acoustic Energy by Wave Field Extrapolation. A. Theoretical Aspect*, 3rd ed., Elsevier Science, Amsterdam, 1985.
9. D. J. Verschuur, "Surface-Related Multiple Elimination: An Inversion Approach," Ph.D. Thesis, 1991.
10. W. E. A. Rietveld, A. J. Berkhout, and C. P. A. Wapenaar, "Optimum Seismic Illumination of Hydrocarbon Reservoirs," *Geophysics*, Vol. 57, 1992, pp. 1334–1345.
11. D. J. Fyfe and P. G. Kelamis, "Removing Coherent Noise Using the Linear Radon Transform: Expanded Abstracts," 54th Annual Meeting of SEG, 1992, pp. 1465–1468.
12. C. P. A. Wapenaar and A. J. Berkhout, *Elastic Wave Field Extrapolation*, Elsevier Science, New York, *1989.*
13. A. J. Berkhout, "Least-Squares Inverse Filtering and Wavelet Deconvolution," *Geophysics*, Vol. 42, No. 7, pp. 1369–1383.

14. A. Ziolkowski, *Seismic Deconvolution*, IHRDC, Boston, 1984.
15. J. F. Claerbout, *Fundamentals of Seismic Processing*, Blackwell Scientific.
16. R. H. Stolt and A. K. Benson, *Seismic Migration, Theory and Practice*, Geophysical Press, London, 1988.
17. A. J. Berkhout, "A Unified Approach to Acoustical Reflection Imaging, Part I: The Forward Model," *J. Acoust. Soc. Am.*, Vol. 93, No. 4, Pt. 1, 1993, pp. 2005–2017.
18. K. Aki and P. G. Richards, *Quantitative Seismology*, Freeman, San Francisco, 1980.
19. A. Tarantola, *Inverse Problem Theory, Method for Data Fitting and Model Parameter Estimation*, Elsevier Science, Amsterdam, 1987.
20. G. J. Lörtzer and A. J. Berkhout, "An Integrated Approach to Lithologic Inversion. Part I: Theory, *Geophysics*, Vol. 57, 1992, pp. 233–245.

70

EFFECTS OF HIGH-INTENSITY SOUND ON STRUCTURES

B. L. CLARKSON

1 INTRODUCTION

High-intensity acoustic pressures can cause strong vibrations of structures on which they impinge. These vibrations may lead to excessive wear, equipment malfunction, or even structural fatigue. The problems tend to be most severe in lightweight structures such as aircraft skin panels subjected to intense noise from jet or rocket engines. In these cases, the noise exposure is most intense at take-off and lasts for some seconds before reducing as the vehicle gathers forward speed. Some examples of early failures and the first attempts at a design procedure are given by Belcher.[1] His later work[2] gives a comprehensive review of the design process and the testing associated with an aircraft development program. In some isolated cases, there has been excessive vibration and subsequent structural failure in process plants where high internal noise levels have been generated in a gas or fluid by discontinuities in pipes or by pumps or fans.[3] Very short duration sounds such as those associated with explosions or the boom produced by aircraft flying at supersonic speed can also cause damage to windows and cracking of plaster on walls of buildings.[4]

The first part of this chapter describes the sound pressure levels produced by high-intensity noise sources such as jet and rocket engines, turbulent boundary layers, cavities, and oscillating shock waves. Design data for pressure levels and spatial distribution are given. The response of typical lightweight structures to these pressures is discussed and a design procedure outlined. The fatigue life of such structures subjected to high frequency, low stress level loadings is then discussed and random fatigue curves given. Finally, the effect of sonic booms on buildings is considered. The vibration produced in windows and walls by the sonic boom is compared with environmental effects such as severe storms, heavy traffic, and so forth.

The discussion in this chapter will concentrate on aerospace vehicles as it is for these structures that the greatest information is available and most comprehensive studies have been made.

2 THE ACOUSTIC LOADING ON AIR VEHICLES

The main feature of acoustic pressures that distinguishes them from other structural loads is that they have higher frequency components that may match many natural modes of vibration of the structure and they are also distributed over the whole surface of the structure and penetrate the interior. The response of any part of a structure to such loads is dependent on the spatial distribution of the loads, their frequency content, and time duration. A general guide for the estimation of these structural loads has been produced by Ungar, Wilby, and Bliss[5] and Eldred.[6]

2.1 Jet and Rocket Noise

As the high-velocity efflux of an aircraft jet engine exhausts into the surrounding air, the turbulent mixing of the gases creates a broadband sound source. See also Part III. Early theoretical work by Lighthill[7,8] showed that the sound intensity levels should be approximately proportional to the velocity of the jet raised to the power of 8, and this was confirmed in early experiments by

---

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker
Copyright © 1997 John Wiley & Sons, Inc.

many workers. The acoustic pressure is thus proportional to V^4. In the case of rocket motors the efflux has a much higher velocity and the increase in acoustic pressure is closer to V^3. A recent review of the current state of knowledge has been produced by Hubbard.[9]

Spatial Distribution of Sound Pressure The mixing region of the jet produces a highly directional sound source. A typical contour pattern is shown in Fig. 1 from which it can be seen that the angle of maximum sound pressure is about 30° to the jet axis. When jets are installed in aircraft, the effect of this directionality of the sound field and the geometry of the structure is to produce localized regions of high sound pressure where the structure intersects the 30° cone. For rocket engines the line of maximum sound pressure is closer to 50° to the axis of the efflux.

Frequency Content of Jet Noise The characteristic of the turbulent zone, which forms the source of jet noise, is such that the wide range in size of mixing eddies produces a sound output that contains a wide range of frequencies. A typical spectrum of the pressure at a point in the near field is broadband with a peak in the frequency region 100–1000 Hz as shown (in nondimensional form) in Fig. 2. At a point in the near field changes in jet velocity change the shape of the spectrum because as the velocity increases the level of the peak increases and it also moves upward in frequency. Along a line parallel to the jet axis at any one jet velocity the frequency of the peak decreases as the distance from the jet nozzle increases. The frequency at which the peak occurs in the spectrum is inversely proportional to the nozzle diameter. Thus in the case of large rocket engines, the peak occurs in the lower frequency region below 100 Hz.

Spatial Correlation of Sound Pressures A full description of the vibration excitation potential of the noise field requires a knowledge of the spatial "phase" relationships of the pressures over the surface of the vibrating structure. However, this effect is usually of second order in response calculations and is neglected in the design methods.

Design Data for Jet Noise In the initial design for an aerospace structure, the noise pressure field distribution and spectra should be obtained from the engine manufacturer. However, if this is not possible a number

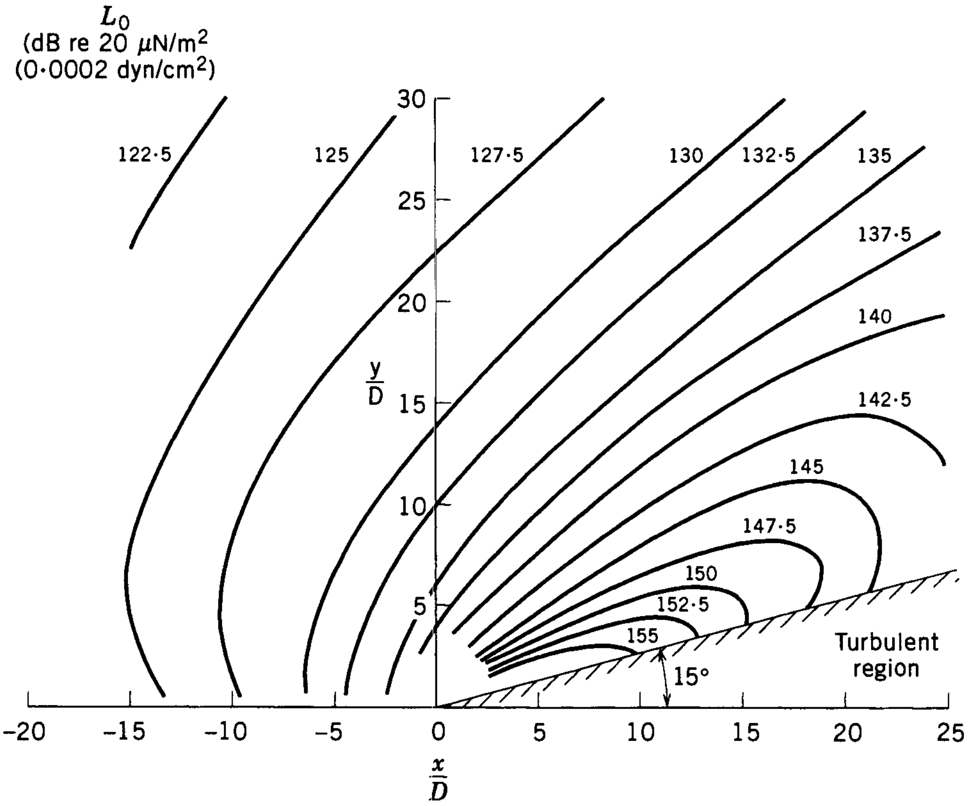


Fig. 1 Near-field sound pressure levels close to a jet exhaust. (From ESDU data sheet 72002. Reproduced with permission of ESDU, London.) D = nozzle diameter.

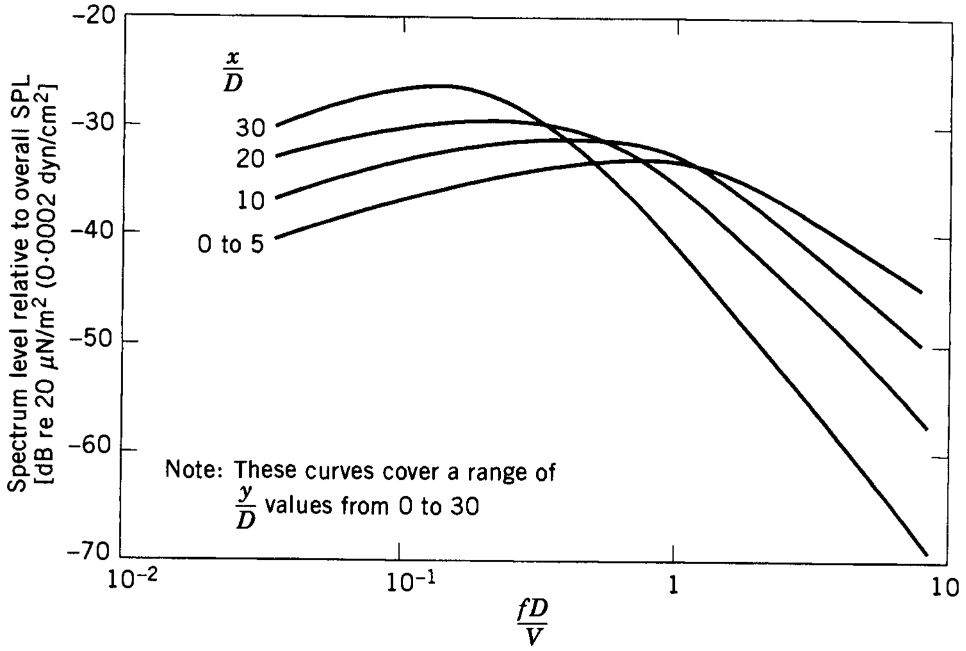


Fig. 2 Spectra of jet noise at four positions downstream of the nozzle. (From ESDU data sheet 72002. Reproduced with permission of ESDU, London.) D = nozzle diameter, V = jet velocity, x = distance downstream of nozzle.

of design procedures has been proposed. The Engineering Sciences Data Unit[10] has produced a nondimensional form of the noise field around a jet as shown in Figs. 1 and 2 from a study of the estimation methods being used by industry and a comparison with the many full-scale measurements around jet engines. This method gives the closest agreement with all the data available, and the accuracy of prediction for a new design is likely to be within ±5 dB. The reference spatial distribution (Fig. 1) enables the level at a point in the near field to be calculated from a knowledge of the nozzle diameter D, jet velocity V, and temperature. A second reference curve (Fig. 2) gives the spectrum at the point to be estimated and is accompanied by a computer program.

Time History The pressure on the surface of the structure is approximately proportional to the jet velocity raised to the power 4, but the velocity is that of the jet relative to the surrounding air with which it is mixing. Thus, as the vehicle begins to move forward at take-off, the relative velocity decreases and hence the sound pressure decreases rapidly. The result is that the maximum pressure occurs at the start of take-off and only lasts for a short time (typically 20–30 s) before falling to a much lower level. In rockets after lift-off the level falls and then rises again at around Mach1 and also at the point of maximum dynamic pressure. These latter levels are usually a little lower than those at liftoff.

Ground Reflections and Thermal Interactions In vertical take-off aircraft, the jet impinges on the runway surface at take-off. There is then a significant reflection effect that must be added to the estimates of pressure loading on the structure and also a heating of the structure from the hot gases of the jet. Theoretical and full-scale experimental work on this effect is described by Scholton.[11] The rate at which the temperature rises will not be as high as the rate at which the noise pressures build up and so the relative time histories of the two effects must be considered in assessing the combined effects on the structure.

2.2 Other Broadband Pressure Loads

There are many other aerodynamic sources that produce broadband pressure fluctuations on regions of aerospace structures. The most important of these are turbulent boundary layers, separated flows, and oscillating shocks. Figure 3 shows the results of measurements made by Coe and Chyu[12] of the pressure spectra caused by these sources. The typical spectrum in a turbulent boundary layer with no static pressure gradient is relatively flat, as shown in the lowest of these curves. The early definitive measurements were made by Bull[13] in a subsonic boundary layer and confirmed that the overall pressure level was approximately equal to 0.006 q_∞ where q_∞ is the free stream dynamic pressure. Subsequent measurements

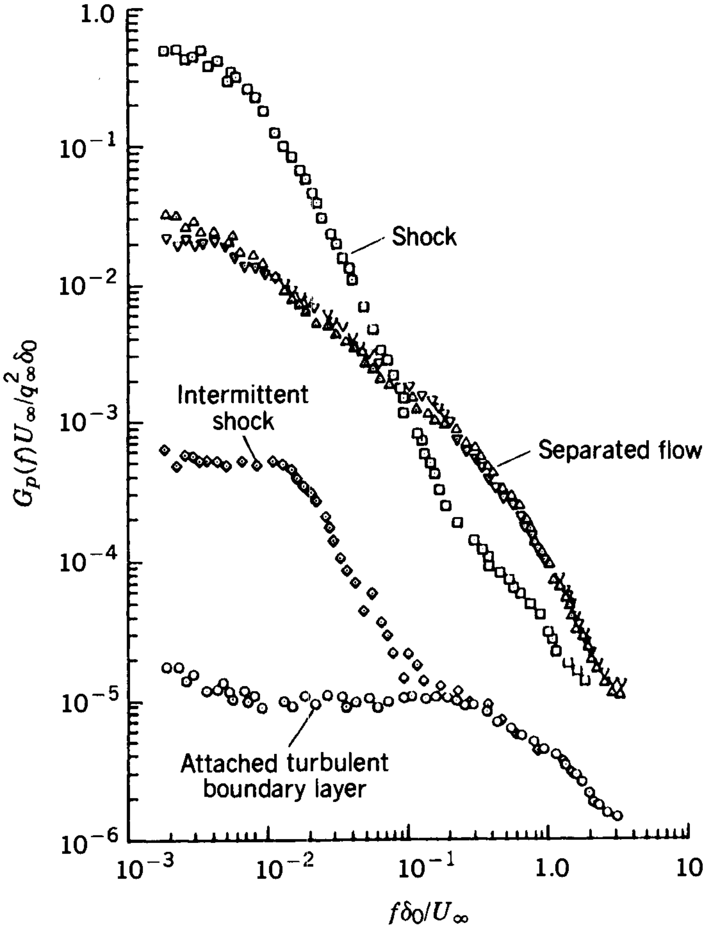


Fig. 3 Typical power spectra of pressure fluctuations underlying supersonic flow (Ref. 12). $G_{\tilde{p}}(f)$ = power spectral-density of pressures, U_∞ = mean free flow velocity, $q_\infty = \frac{1}{2}p_\infty u_\infty^2$ = dynamic head, p_∞ = density of median, δ_0 = boundary layer thickness.

in supersonic flow reported, for example, by Laganelli and Martellucci[14] indicated that the overall levels fall to about $0.003q_\infty$ at a Mach number of 3.5. The frequency spectrum is relatively flat until roll-off occurs at a point determined by the boundary layer thickness δ_0 and the flow velocity u_∞. The roll-off frequency is proportional to the velocity and inversely proportional to the boundary layer thickness. From a structural point of view, the spectral density in the 200–1000-Hz region is usually the most important excitation mechanism. For a given vehicle at high speed the overall pressure level is approximately constant along the length. However, the boundary layer thickness increases toward the rear of the vehicle and consequently the roll-off frequency decreases and the spectral density increases. Thus, the damaging part of the spectrum is increased in level toward the rear.

Separated flow such as occurs behind wedges, discontinuities and at the mouth of cavities causes high fluctuations in the local region. Few measurements are available, but one set is also shown in Fig. 3. In this case the spectrum falls off steadily with frequency, but the spectral density in the lower frequency region is two or three orders of magnitude higher than that produced by a typical turbulent boundary layer. This may be sufficient to cause local failures. Reference 5 contains a good data summary of cavity noise measurements.

Oscillating shocks can also cause very high pressures on very localized areas of structure. A typical spectrum has been obtained by Coe[12] and is also shown in Fig. 3. A further increase in the spectral density in the lower frequency range is observed.

Two further examples of feedback mechanisms may be important in some designs. Shock cell noise resulting from eddy/shock front interaction in a choked jet is described by Hay and Rose.[15] In-flight pressures similar to those at-take-off can be produced on structures close to the jet. In the second example a feedback mechanism can occur when a flap is lowered into a jet flow as occurs in some high lift configurations.[16] The basic mechanism behind these phenomena is a flow instability or an acoustic-flow feedback loop. The mechanism is understood in physical terms, and hence the frequency of occurrence can be estimated but the magnitude of the disturbance cannot be calculated theoretically. Thus we must rely on measurements and empirical formulas to produce estimates for new designs.

2.3 Propeller Noise

The proposed high-speed propellers for new designs of aircraft (prop fans, unducted fans, ultra-high bypass engines, advanced turboprops, etc.) generate high sound pressure on the fuselage side during cruise. This could become a fatigue problem because of the long duration. Some flight test data for such a design are given in Ref. 17.

3 RESPONSE OF STRUCTURES

3.1 Metallic Structures

The high-intensity acoustic noise sources cause the structure in the region of the most intense pressures to vibrate at high frequency. This vibration can cause low-amplitude oscillating stresses that can build up a sufficiently large number of stress reversals to cause fatigue crack initiation at points of stress concentration. The majority of the examples of failure of aircraft structures has been confined to skin failures along rivet lines, skin debonding in sandwich panels, and failures in internal attachment structure. Light-weight internal support structure such as frames, ribs, and cleats is prone to failures at the bend radius, which are difficult to detect from the outside. Many studies have been made of the response and subsequent fatigue of whole sections of aircraft structure, groups of panels, and single panels (see, e.g., Lin[18] and Clarkson and Ford[19]). Design procedures have been developed from these results and from theoretical studies of the response of simplified structural models.[20,21]

Although the response of a typical aircraft structure to broad-band random pressures is very complicated, some trends in behavior can be observed. For relatively large panels the response is usually dominated by the first mode of vibration.[20] In rows of smaller interconnected panels, such as occur in fuselage sides, the response is more complicated with adjacent panels influencing each other and responding as a group. However, even in this situation, the linkage between the response of several panels in an array is usually weakened by structural discontinuities and variation in panel dimensions that occur in real aircraft. Typical response spectra are given in Fig. 4.

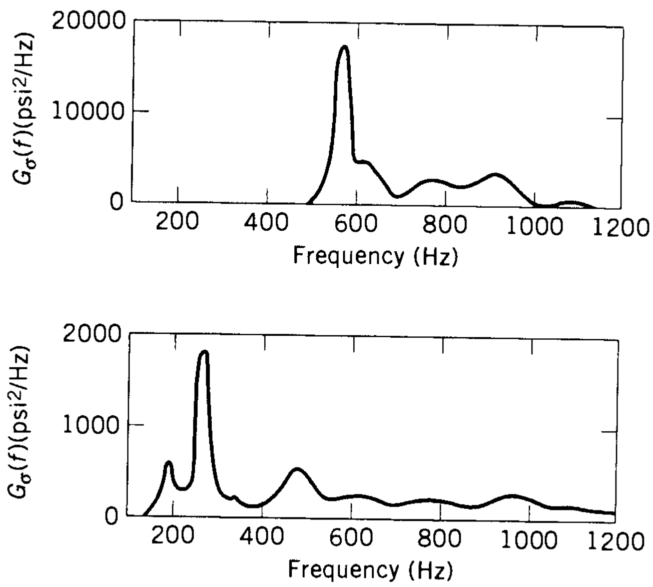


Fig. 4 Early stress spectra for fuselage skin side panels (Ref. 19). Reproduced from "The Aeronautical Journal" by kind permission of the Royal Aeronautical Society.

A design procedure based on the theory of the response of a single-degree-of-freedom system to a random force having a constant spectral density (the white noise model) was first derived by Miles[22] and forms the basis of most estimation procedures. In this model the structure is represented by a single panel, and the noise excitation is assumed to take the form of a uniform pressure field having a constant spectral density, say $G_p(f_n)$, at the fundamental frequency of the panel, f_n. The spectral density of the response can then be integrated over the frequency domain to produce the now much used expression for the mean-square stress response:

$$\sigma^2(t) = K\frac{\pi}{4\eta} f_n G_p(f_n)\left(\frac{\sigma_0}{f_0}\right)^2 \qquad (1)$$

where η is the panel's loss factor and σ_0 is the static stress produced by a uniform pressure F_0 over the surface. The factor K has been added to allow the simple formula ($K = 1$) to be used empirically to fit-experimental data. Usually the stress of interest is that at the rivet or attachment line and so σ_0 would also be calculated at this point. This relationship shows that the rms level of the response is inversely proportional to the damping. A survey of damping measurements for typical aircraft structures has been produced by ESDU.[10] This shows that the average damping of a typical metallic structure is of the order of 1% of crictical damping (or $\eta = 0.02$). Reductions in stress can be achieved by adding damping to the structure in the form of viscoelastic interlayers or coatings. A comprehensive review of damping mechanisms and methods of increasing the damping of aerospace structures has been produced by Soovere.[23] See also Chapter 71.

The most general design method has been developed by the Engineering Sciences Data Unit[10] based on the simple formula given by Eq. (1). In the ESDU data sheets, procedures are given for the estimation of each of the parameters in the equation. Natural frequency curves are given for a range of plate configurations and now include fiber-reinforced and honeycomb plates as well as flat and curved aluminum. The damping for typical structures is given in another set as mentioned above, and the spectral density of the forcing pressure is estimated from the sound field measurements or estimates described in an earlier section. Methods of calculating σ_0/F_0 for the plate edges are also given. The values of the parameters given in the data sheets have been derived from an analysis of the test results available, including unpublished company results. Thus for each parameter in the equation, the designer can use a data sheet based on a wide range of tests, or he or she can use directly measured quantities for one or more of the parameters if they are available for his or her own design. This gives a flexible design procedure that can readily be adapted to include the latest information.

The very early design work was based on test results and the simple Eq. (1) but was presented in the form of an easy-to-use nomograph (see, e.g., Belcher et al.[1] and Ballentine et al.[24]). The nomograph shown in Fig. 5 is for the choice of skin thickness in a flat stiffened skin design where the individual panels have the dimension 7.2 by 4.8 in. To use the figure one takes a series of straight-line paths between reflecting lines as shown. One first decides the number of stress reversals that the structure must endure in its lifetime. In this example we choose 6 × 10^8 and start from this point on the baseline. At the first series of reflecting lines a selection of confidence limits is made to allow for the scatter in fatigue results. If one is going to test the structure, a lower limit might be acceptable or alternatively if one wishes to be conservative a higher limit (lower stress) is more appropriate.

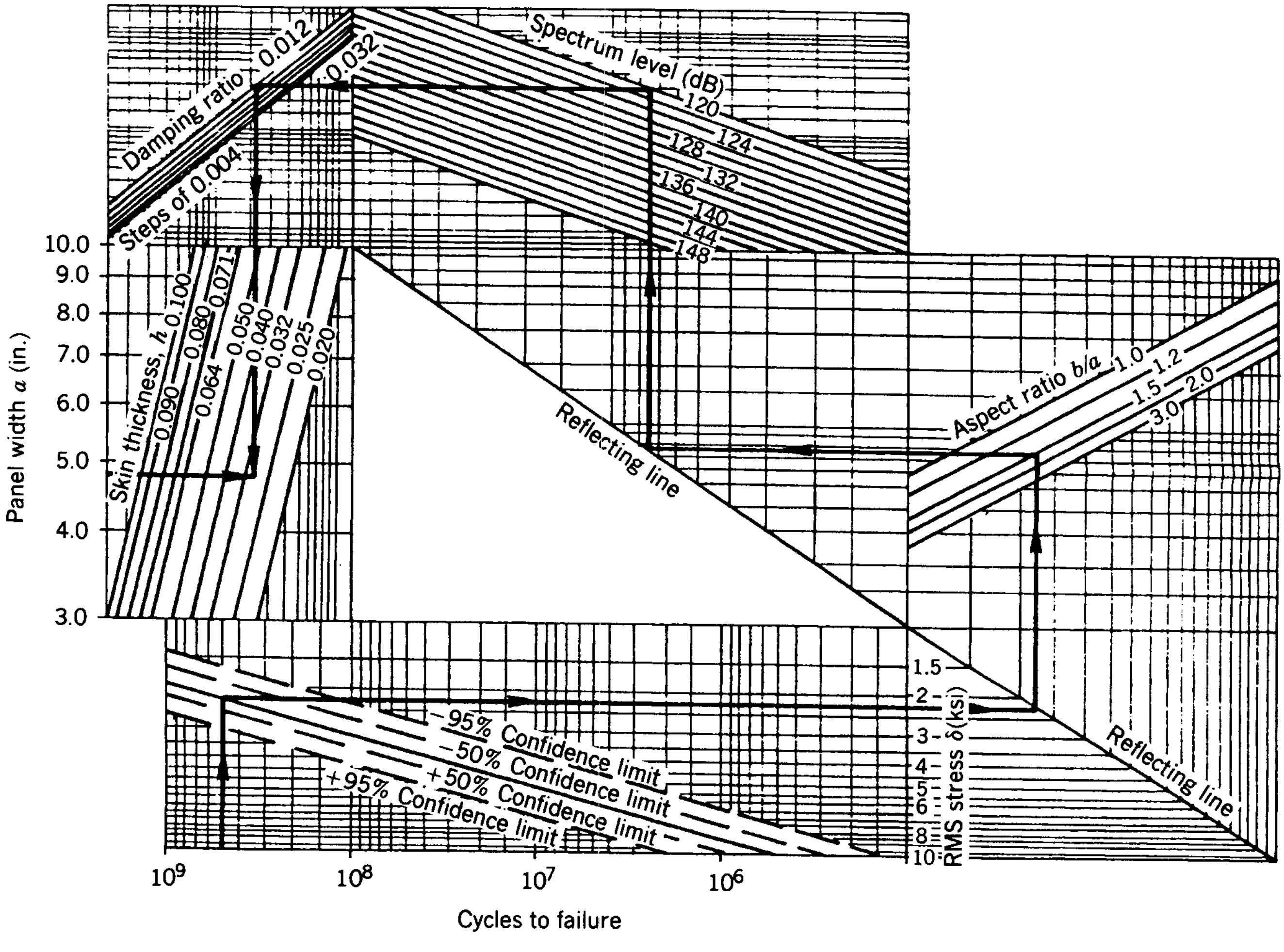


Fig. 5 Nomograph for stress and fatigue life for stiffened panels (Ref. 24).

In this case the mean line has been selected. After passing the corner of the next reflecting line we come to the choice of aspect ratio of the panel (1.5 in this case). From here we go on to the spectrum level of the acoustic pressure field, which will have been obtained from one of the estimation procedures described in the previous section. Next we come to the choice of damping ratio (0.012 in. this case). We now come to the final step, which is to determine the skin thickness. Drawing a horizontal line from the known panel width (4.8 in.), we intersect the line reflected from the damping ratio at a thickness of 0.3 in. If this is not acceptable from other considerations, the process could be repeated by a more radical redesign such as changing the panel width. The next increase in sophistication in such a nomograph is to add a sector for panel curvature. The full range of nomographs for the different types of structure tested is given by Rudder and Plumblee.[25] The designer can now use one chart to complete the whole process. The results from this method tend to be more accurate than the ESDU method when the new structure is close to that used in the tests, which formed the basis of the nomograph. They are not as good when the structure deviates from that which was tested.

Equation (1) applies strictly only to the case where one mode of response is predominant, such as in the response of large flat panels and honeycomb sections. Where there is significant coupling of the surface panels with adjacent panels in a fuselage side or with ribs in a control surface, many more modes of vibration are excited to give significant response levels and contribute to the overall stress. However, attempts to produce a more accurate method have not yet been successful. Arcas[26] has produced a modification of the ESDU method and compared actual experimental results from the two together with Ballantine's nomograph. These comparisons show that there is considerable scatter in the results.

3.2 Honeycomb and Composite Structure

Some of the earliest failures occurred in simple aluminum skin/rib designs for flaps close to or impinging on jet effluxes. The next step was to stiffen up the skin by replacing the aluminum skin with a honeycomb plate. Such structures were found to have superior resistance to acoustic fatigue and careful design of the edge close-out section led to very efficient designs. Tests on honeycomb panels provided data for this type of structure to be included in the ESDU data sheets and also the series of design nomographs.[25]

Because of their high stiffness-to-weight ratio, carbon fiber-reinforced plate structures are now being used more extensively in aerospace structures. Such composite plates also show good resistance to acoustic loading.[27] Tests on single panels again show a predominantly single-mode response and estimates using the simple theory show good agreement with test results (Fig. 6). In this case the structure is driven into a nonlinear response range at the upper limit of the design stresses. This nonlinearity is evident in the way the actual stress begins to fall below that estimated at higher stress levels. Carbon fiber-reinforced plates are also being used as face sheets for honeycomb panels. These can be designed to give very good resistance to acoustic fatigue.

4 FATIGUE

The main characteristic of acoustic loading is its high frequency compared with other fatigue-inducing loads on an aircraft. Although the high noise levels at take-off only last for about 20 s, the high frequency of the structural response (100–500 Hz) means that there can be 10,000 stress reversals at each take-off. thus in the lifetime of an aircraft there will be 10^8–10^{10} reversals. This is beyond the range of traditional fatigue data, and therefore special tests have been developed to provide appropriate results. To provide data for designers, tests on typical joints (panel edge mountings) have been made using a random wave form of the modulated sine wave type and endurance results plotted against the root-mean-square (rms) stress level. A typical curve is shown in Fig. 7. A comprehensive series of curves for aluminum and titanium alloys and carbon fiber-reinforced plastic is given in the ESDU data sheets.

5 SONIC BOOM

When aircraft began to fly supersonically, there were some well-publicized cases of windows being broken by the sonic boom. This led to a study of the effect of the atmosphere on the propagation of the boom, the pressure signature on the ground produced by different sizes of aircraft, and on the effect of these pressures on buildings. These studies reached a climax in the run up to the commercial operation of *Concorde*.[28] Finally it was the fear of damage from the sonic boom that restricted commercial supersonic flights to routes over water. Little new work was done after that stage, but there is now a renewed interest in the subject as several aircraft manufacturing consortia are considering designs for a second-generation supersonic transport.

5.1 Pressure Signature

A good review of the pressure loading produced by aircraft flying supersonically is given by Maglieri and

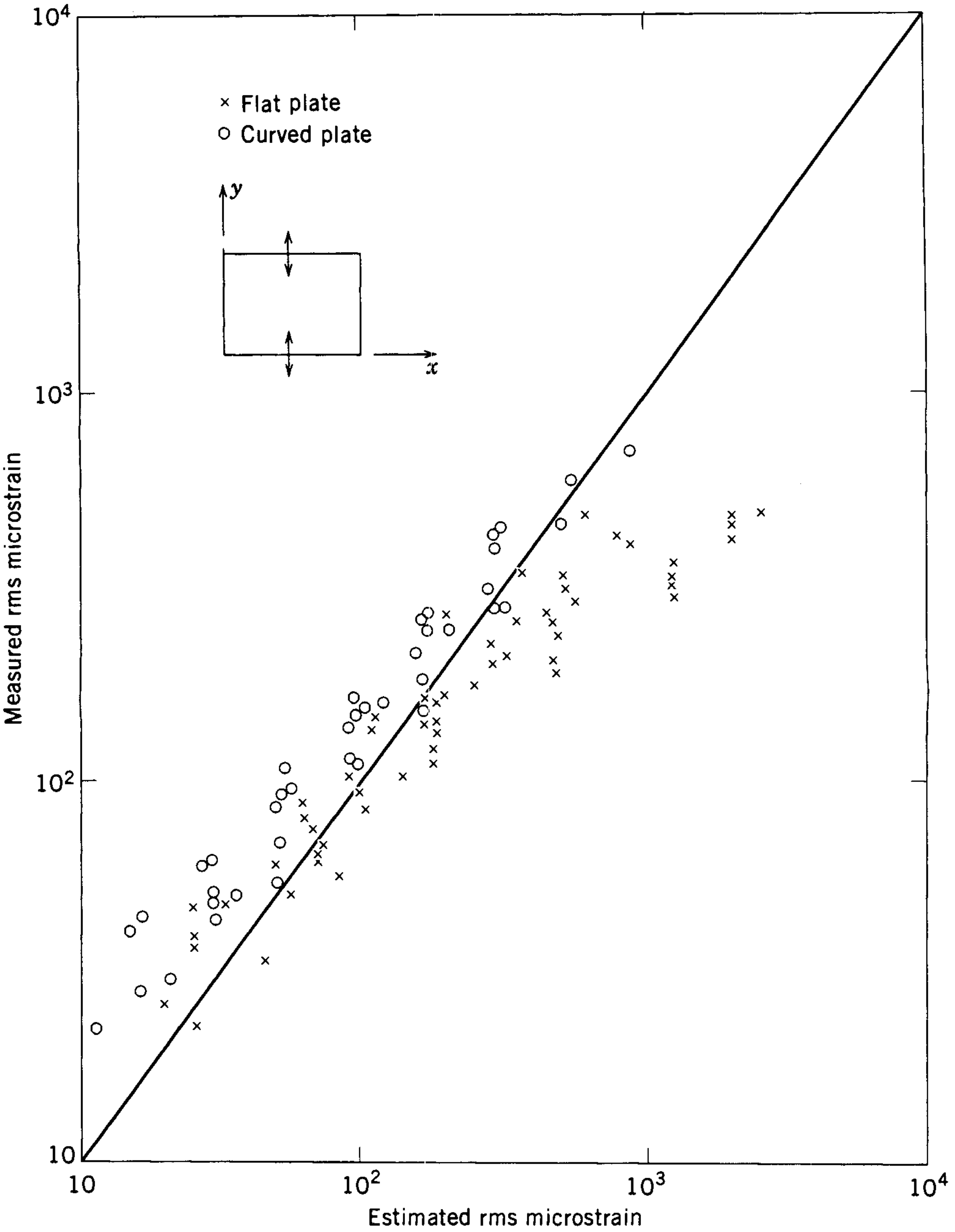


Fig. 6 Comparison of measured and estimated edge strains in carbon fiber-reinforced plastic plates. (From ESDU data sheet 84008. Reproduced with permission of ESDU, London.)

Plotkin.[28] The pressure wave on the ground associated with the sonic boom is of the N form, as shown in the *top trace* of Fig. 8. This is a relatively pure-wave form such as would be measured in unobstructed flat countryside. In a built-up area, reflections from nearby structures and diffraction of the wave as it passes over a building will distort the simple N shape. Aircraft maneuvers and major atmospheric effects such as temperature inversions can cause focusing of the rays and amplification of the bang. The magnitude of the peak pressure is dependent on the aircraft weight and altitude. The duration of the N is a function of the aircraft size. (See also Chapter 31.)

5.2 Response of Structures to Sonic Booms

The response time history of typical structural elements of timber-framed houses are also shown in Fig. 8. This shows that the initial impulse and also the final rapid

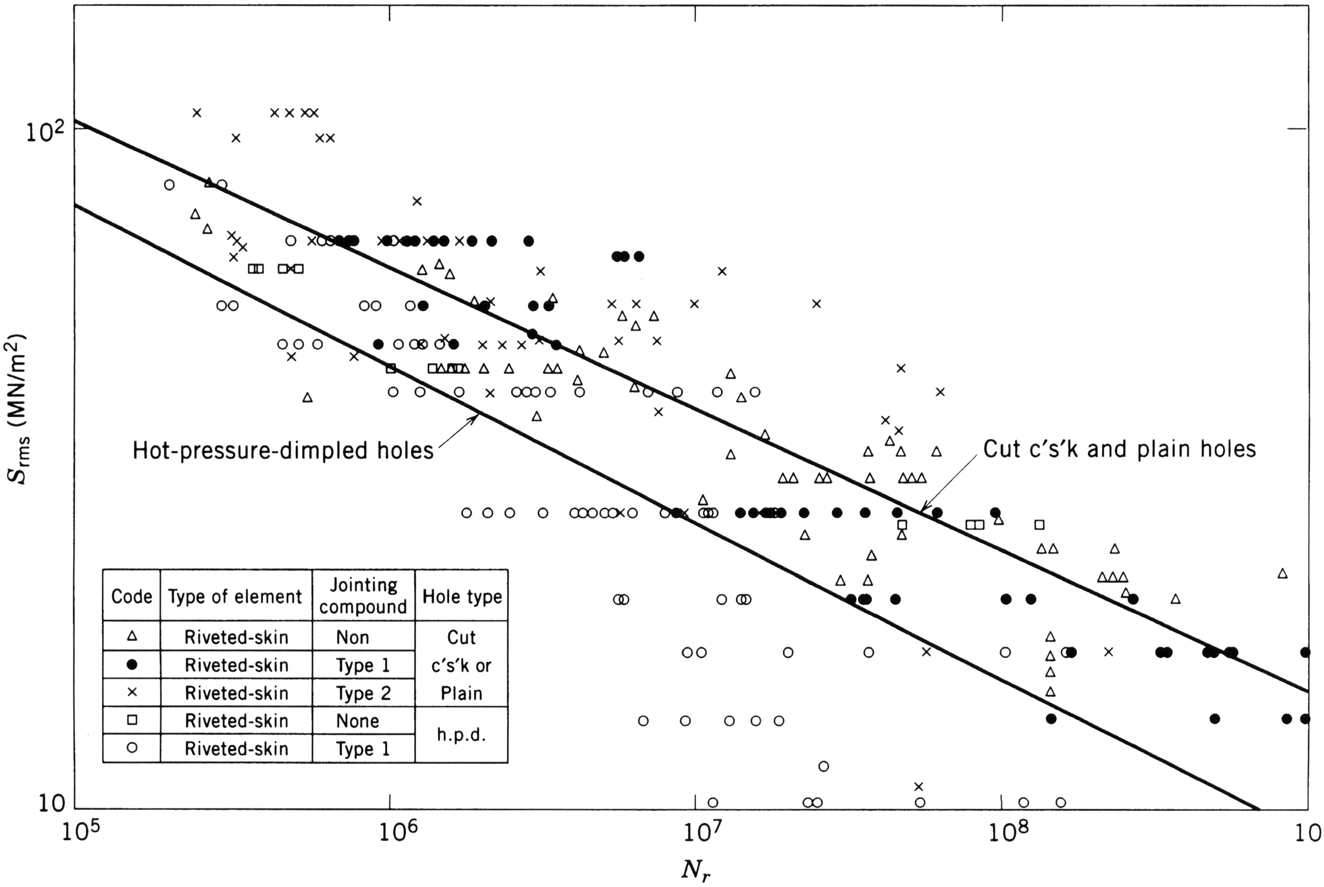


Fig. 7 Endurance data for riveted-skin joints subjected to narrow-band random loading. (From ESDU data sheet 72015. Reproduced with permission of ESDU, London.) S_{rms} = rms value of narrow-band random stress, N_r = number of cycles to failure.

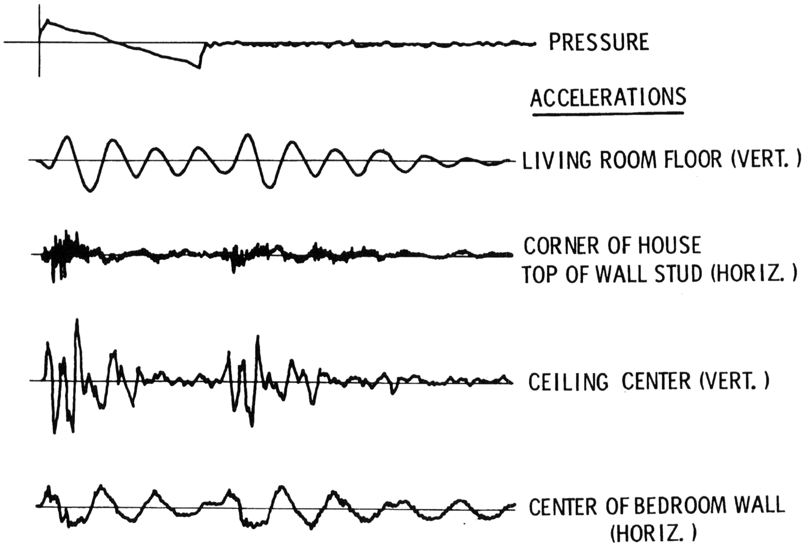


Fig. 8 Time histories of forcing pressure and acceleration response of parts of a typical American home.

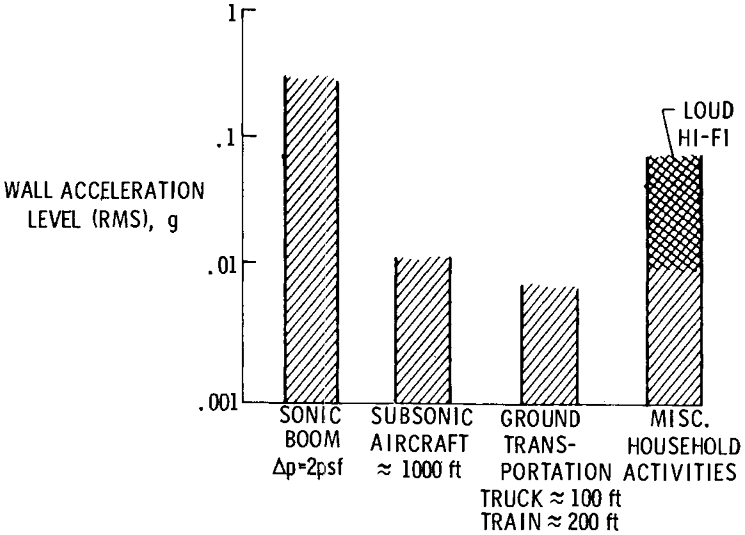


Fig. 9 Maximum wall acceleration levels produced by various disturbances (Ref. 30).

return to zero overpressure set up pulses of vibration in the structure. This figure shows that the walls, floors, and ceilings have responses that exhibit different predominant natural frequencies and that the damping is low. Although a multimodal response is induced by the boom, calculations based on the response of each element in its first mode of vibration should give a reasonable estimate of the maximum level.

Using the single-degree-of-freedom model and an idealized N-shaped waveform, the maximum response can be calculated and expressed as a proportion of the maximum response that would have been produced by a static load equal in magnitude to the peak pressure in the N wave. This can be written in the form of a dynamic amplification factor (DAF). Calculations by Cheng and Benveniste[29] shows that for likely designs of supersonic transport, the dynamic amplification for all the building components will be in the range 1.5–2.0. There have been several attempts to refine the calculation procedures by using a much more sophisticated model of the structure. Finite-element methods have been used, but no greater accuracy has been achieved than is available from the simple DAF design curves. The great variation in structural detail design and the maintenance condition of buildings means that it is hardly worthwhile trying to refine a computational model.

Acknowledging that it is not possible to make very accurate predictions of response, comparisons with other common causes of structural vibration have been made in order to help assess the likelihood of damage. One such comparison for wall accelerations is given by Cardin[30] in Fig. 9. This shows that the levels produced by the sonic boom are about an order of magnitude higher than those for common environmental events.

5.3 Possible Damage to Historic Buildings

Many historic buildings such as old churches and manor houses are in a rather fragile state. Stained glass windows and decorative plaster work are susceptible to damage from blast loading. A series of tests was made in Britain

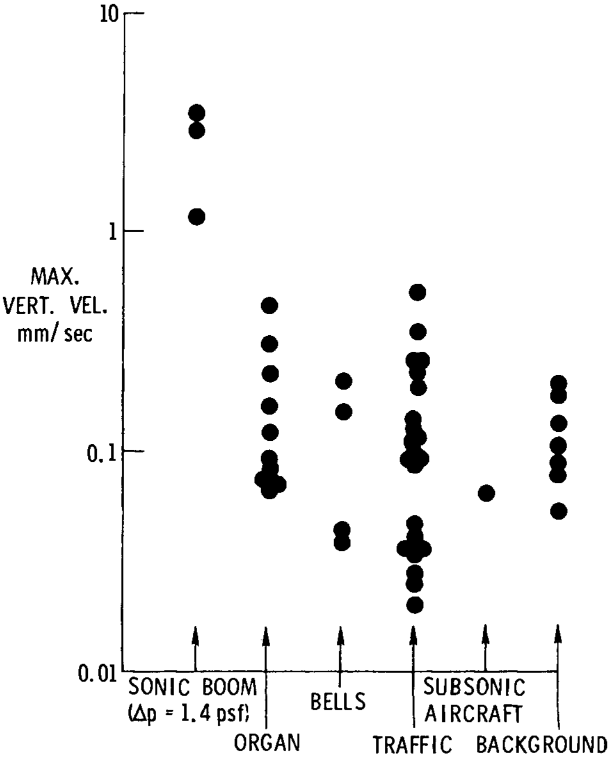


Fig. 10 Response of cathedral vaulting to transient pressures (Ref. 4).

to measure the response of cathedral windows and vaulting to impulsive pressure loading arising from sonic booms, traffic, noise, and other environmental noises. The results for a vault are shown in Fig. 10. This figure shows that the levels produced by the sonic boom are about an order of magnitude greater than those produced by the commoner environmental disturbances.

5.4 Damage Levels

Safety criteria operated by the U.S. Department of Mines for blasting operations suggest that damage is likely to occur to buildings when peak velocity exceeds 125 mm/s. This compares with measured values on the cathedral vault of about 6 mm/s. This suggests that there is an order of magnitude difference between sonic boom levels and damage level. However, because of the great variability in response of old structures, such a difference may be required to ensure that no damage is sustained.

REFERENCES

1. P. M. Belcher, J. D. Dyke, and A. L. Eshleman, "Development of Aircraft Structure to Withstand Acoustic Loads," *Aerospace Eng.*, Vol. 18, No. 6, 1959.
2. P. M. Belcher, "Selected Topics from the Structural Acoustics Program for the B1 Aircraft," *Shock Vib. Bull.*, Vol. 49, Part 3, 1979, pp. 55–68.
3. W. Rizk and D. F. Seymore, "Investigations into the Failure of Gas Circulators and Circuit Components at Hinkley Point Nuclear Power Station," *Proc. I. Mech. E.*, Vol. 179, No. 1, 1964, pp. 627–703.
4. B. L. Clarkson and W. H. Mayes, "Sonic-Boom-Induced Building Structure Responses Including Damage," *ASA 2nd Sonic Boom Symposium*, Houston, Texas, Nov. 1970.
5. E. E. Ungar, J. F. Wilby, and D. B. Bliss, "A Guide for Estimation of Aeroacoustic Loads on Flight Vehicle Surfaces," AF FDL-TR 76-91, 1976.
6. K. M. Eldred, "Acoustic Loads Generated by the Propulsion System," NASA SP-8072, 1971.
7. M. J. Lighthill, "On Sound Generated Aerodynamically. I. General Theory," *Proc. Roy. Soc.*, Vol. A211, 1952, pp. 564–587.
8. M. J. Lighthill, "On Sound Generated Aerodynamically. II. Turbulence as a Source of Sound," *Proc. Roy. Soc.*, Vol. A222, 1953, pp. 1–32.
9. H. H. Hubbard (Ed.), "Aeroacoustics of Flight Vehicles: Theory and Practice," NASA RP-1258, 1991.
10. *ESDU Engineering Data; Acoustic Fatigue Series*, Vol. 1–6, ESDU International, London,
11. R. Scholton, "Influence of the Ground on the Near Field Noise Levels of a Jet Supported V/STOL," AGARD CP 113, 1973.
12. C. F. Coe and W. J. Chyu, "Pressure Fluctuation Inputs and Response of Panels Underlying Attached and Separated Supersonic Turbulent-Boundary Layers," Section 5 in AGARD CP-113, 1973.
13. M. K. Bull, "Wall Pressure Fluctuations Associated with Sub-sonic Turbulent Boundary-layer Flow," *J. Fluid Mech*, Vol. 28, Pt 4, 1967, p. 719.
14. A. L. Laganelli and A. Martellucci, "Wall Pressure Fluctuations in Attached Boundary-layer Flow," *AIAA J.*, Vol. 21, 1983, pp. 495–502.
15. J. A. Hay and E. G. Rose, "In Flight Shock Cell Noise," *J. Sound Vib.*, Vol. 11, 1970, pp. 411–420.
16. M. R. Fink, "Propulsive Lift Noise," NASA RP-1258, 1991, pp. 449–481.
17. M. A. Simpson, P. M. Druez, A. J. Kimbrough, M. P. Brock, P. L. Bruge, G. P. Mathur, M. R. Cannon, and B. N. Tran, "UHB Demonstrator Interior Noise Control Flight Tests and Analysis," NASA Contractor Report 181897, 1989.
18. Y. K. Lin, "Stresses in Continuous Skin Stiffener Panels under Random Loading," *J. Aerospace Sci.*, Vol. 29, 1962, pp. 67–75.
19. B. L. Clarkson and R. D. Ford, "The Response of a Typical Aircraft Structure to Jet Noise," *J. Roy. Aero. Soc.*, Vol. 66, 1962, pp. 1–40.
20. B. L. Clarkson, "The Design of Structures to Resist Jet Noise Fatigue," *J. Roy. Aero. Soc.*, Vol. 66, 1962, pp. 603–616.
21. B. L. Clarkson, "Stresses in Skin Panels Subjected to Random Acoustic Loading," *J. Roy. Aero. Soc.*, Vol. 72, 1968, pp. 1000–1010.
22. J. W. Miles, "On Structural Fatigue under Random Loading," *J. Aero. Soc.*, Vol. 21, No. 11, 1954, pp. 753–762.
23. J. Soovere and M. L. Drake, "Aerospace Structures Technology Damping Design Guide," AFWAL TR-84-3089, 1985.
24. J. R. Ballentine, F. F. Rudder, Jr., J. T. Mathis, and H. E. Plumblee, Jr., "Refinement of Sonic Fatigue Structural Design Criteria," AFFDL TR 67-156, 1968.
25. F. F. Rudder and H. E. Plumblee, "Sonic Fatigue Design Guide for Military Aircraft," AFFDL TR 74-113, 1974.
26. N. R. Arcas, "Prediction of Stress and Fatigue Life of Acoustically Excited Aircraft Structure," *Shock Vib. Bull.*, Vol. 39, Part 2, 1969.
27. I. Holehouse, "Sonic Fatigue Design Techniques for Advanced Composite Aircraft Structures," AFWAL TR 80-3019, 1980.
28. D. J. Maglieri and K. J. Plotkin, "Sonic Boom," NASA RP-1258, 1991, pp. 519–561.
29. D. H. Cheng and J. E. Benveniste, "Dynamic Response of Structural Elements Exposed to Sonic Booms," NASA CR 1281, 1969.
30. H. D. Carden, D. S. Findley, and W. H. Mayes, "Building Vibrations Due to Aircraft Noise and Sonic Boom Excitation," *ASME Symposium on Machinery Noise*, Los Angeles, 1969.

71

VIBRATION ISOLATION AND DAMPING

ERIC E. UNGAR

1 INTRODUCTION

Isolation and structural damping constitute the two most widely applicable means for the control of vibration or structure-borne sound, particularly in the audio frequency range. This chapter delineates the most important considerations pertaining to passive vibration isolation and damping; active vibration control—achievement of the effects of isolation and damping by means of control systems and sources of energy that are external to the vibrating system—is covered in Chapter 75.

Vibration isolation in essence involves use of a resilient connection between a source of vibration and an item to be protected, so that this item vibrates less than it would if a rigid connection were used. In some typical situations the source consists of a vibrating machine or structure and the item to be protected is a supporting or connecting structure; in other typical situations the source consists of a vibrating structure and the item to be protected is a sensitive piece of equipment supported from that structure. Many salient features of vibration isolation can be analyzed in terms of a simple model consisting of a rigid mass that is connected to a support via a massless linear spring and that is constrained to translate along a single axis. The mass here may represent either the body of a vibration producing machine or a sensitive item to be protected from support vibrations, with the spring representing the isolator. More complex models are needed to address situations where the magnitude of the excitation depends on the motions, where an additional spring–mass system is inserted between the primary one and the support, and/or at comparatively high frequencies where the isolator mass plays a significant role or where the isolated items do not behave as rigid masses. Other complications arise because of non-uniaxial motions and nonlinearities.

Damping refers to the removal of mechanical energy from a vibration of concern. This removal may result from transfer of energy to structural components, fluids, or vibration modes that are not of concern or from conversion of mechanical energy into other forms. Damping affects the decay of free (unforced) vibrations and of those forced responses of systems that are controlled by energy balances. Various measures of damping are in use; most are based on observation of the effects that damping has on the motions of simple systems. Data on the damping inherent in materials cover a wide range, extending from comparatively low damping for high-strength structural materials to very high damping for some viscoelastic materials (typically, plastics or elastomers) with limited strength. Structural components that are acceptably strong and that also exhibit relatively high damping may be obtained by combining high-damping viscoelastic materials with structural materials in the form of added layers or in sandwich arrangements.

2 VIBRATION ISOLATION

2.1 Linear Model with Single Degree of Freedom

The basic idealized system of Fig. 1 represents the simplest model on the basis of which vibration isolation has been analyzed. The combination of the linear spring of stiffness (force/displacement) k and the dashpot with viscous damping coefficient (force/velocity) b represents an isolator (or an array of isolators); the isolator is taken to be massless and the mass m is assumed to be rigid

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker
Copyright © 1997 John Wiley & Sons, Inc.

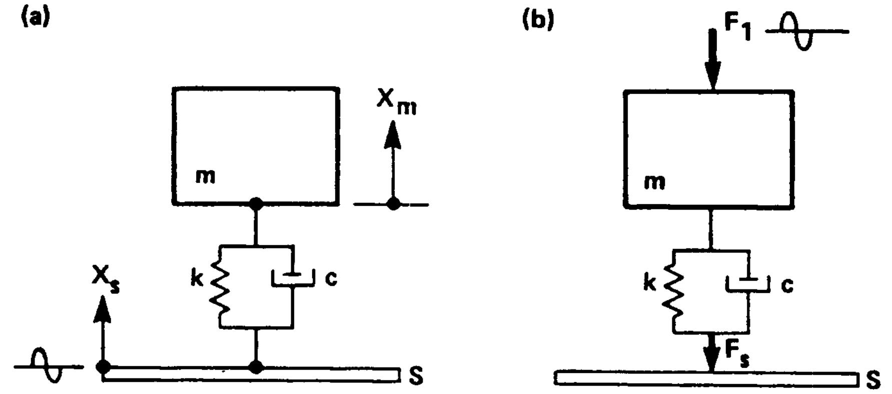


Fig. 1 Mass–spring–dashpot system: (*a*) excited by support motion; (*b*) excited by force acting on mass.

and to be capable of only vertical, purely translational motion. In the case of Fig. 1*a*, where the support is taken to vibrate vertically with a prescribed displacement amplitude X_s at a given frequency, the mass represents an item to be protected from support-induced vibrations. In the case of Fig. 1*b*, where a force of a prescribed amplitude F_1 at a given frequency is taken to act on the mass, the mass represents the frame of a machine from whose vibrations the support is to be protected. The equations of motion for this system are given in Chapter 63 and Ref. 2.

Transmissibility and Isolation Efficiency For the case of support-induced vibrations, the (dimensionless) motion transmissibility $T_v = X_m/X_s$ relates the displacement amplitude X_m of the oscillation of the mass (at the support motion frequency) to the displacement amplitude X_s of the support vibration. For the case where an oscillatory force acts on the mass, the (dimensionless) force transmissibility $T_f = F_s/F_1$ relates the amplitude F_s of the force that the isolator exerts on the support to the amplitude F_1 of the exciting force. If the support in Fig. 1*b* is immobile, then the two transmissibilities are given by the same expression,* namely

$$T_v = T_f = T = \left[\frac{1+\eta^2}{(1-r^2)^2+\eta^2}\right]^{1/2}, \qquad (1)$$

where $r = f/f_n$ is the ratio of the excitation frequency f to the natural frequency f_n of the spring–mass system and η denotes the loss factor of the isolator (see below).

The natural frequency f_n (in hertz) obeys

$$f_n = \frac{1}{2\pi}\sqrt{\frac{k}{m}} = \frac{1}{2\pi}\sqrt{\frac{g}{X_{\mathrm{st}}}} \approx \frac{3.13}{\sqrt{X_{\mathrm{st}}\,(\mathrm{in.})}} = 15.7/\sqrt{X_{\mathrm{st}}\,(\mathrm{mm})} \qquad (2)$$

where g denotes the acceleration of gravity and X_{st} represents the static deflection induced in the isolator (spring) by the mass resting on it.

The (dimensionless) loss factor $\eta = D/2\pi W$ relates the energy D that the isolator dissipates per cycle in the steady state to the timewise maximum strain energy W stored in the isolator in a cycle. Although η may vary with amplitude and frequency in general, it is approximately constant for most metals and in certain limited frequency and temperature ranges also for plastics and elastomers. For viscous damping, in which motion is opposed by a force that is proportional to velocity, the loss factor obeys $\eta = 2rb/b_c$, where b represents the viscous damping coefficient (the constant relating the retarding force to velocity) and $b_c = 2\sqrt{km}$ denotes the so-called critical damping coefficient. See also Ref. 2 and Section 1 in Chapter 63.

Since transmissibility represents the fraction of the disturbing motion or force that is transmitted through the isolator, good isolation (i.e., large vibration reduction) corresponds to small values of transmissibility. A measure of isolation performance that takes on greater values for better isolation is the dimensionless *isolation efficiency* $I = 1 - T$, which represents the fraction of the disturbance that is *not* transmitted. The isolation efficiency generally is expressed in percent. For example, a transmissibility of 0.0073 corresponds to an isolation efficiency of 0.9927, or 99.27%.

From Eq. (1), it is evident that for small damping

---

*Equality of the two transmissibilities, which refer to two different physical situations, is a consequence of the reciprocity principle and holds for any linear uniaxial system, no matter how complex[1] and not just for the single-degree-of-freedom system of Fig. 1.

($\eta^2 \ll 1$), which usually is of greatest practical interest,

$$T \approx 1 \qquad \text{for } r \ll 1, \tag{3a}$$

$$T \approx \frac{1}{\eta} \qquad \text{for } r \approx 1, \tag{3b}$$

$$T \approx \frac{1}{r^2} \text{ or } \frac{2(b/b_c)}{r} \qquad \text{for } r \gg 1, \tag{3c}$$

where the first expression of Eq. (3c) applies for the case of a constant, frequency-independent loss factor, and the second applies for viscous damping, in which the loss factor is proportional to frequency.

In order to obtain good isolation, one generally wants a system that operates in the range where $f \gg f_n$; one thus wants to make the natural frequency f_n as small as possible compared to the excitation frequency f. In view of the various complications that limit the applicability of Eq. (1), as discussed later, transmissibilities that are smaller than $\frac{1}{100}$ over extended frequency ranges can be obtained only rarely with practical conventional simple isolation systems. Reduced transmissibility in the resonance region ($f \approx f_n$) can be achieved only by increasing the damping, that is, the loss factor η.

Effect of Damping. Although Eqs. (1) and (3c) indicate that increased *viscous* damping may result in significant increases in transmissibility at high frequencies, this fact tends to be of little practical importance, not only because the damping of practical isolators generally is relatively small, but also because their damping behavior typically corresponds more nearly to hysteretic damping (as characterized by a frequency-independent loss factor) than to viscous damping.

Machines with Unbalance.[2] In a rotating or reciprocating machine, the dominant excitation force F_u generally is due to unbalance. This force occurs at the frequency f that corresponds to the machine's rotational speed and is given by $F_u = m_u e_u (2\pi f)^2$, where m_u denotes the unbalance mass and e_u its eccentricity (i.e., its distance from the axis of rotation). For such a frequency-dependent force, the force that acts on the structure that supports the machine is given by $F_s = F_u T_F$. At frequencies f that are much greater than f_n, this force approaches $F_s = k e_u m_u / m$. In contrast to the situation where the excitation force is frequency independent, the force transmitted to the support here does not decrease with increasing excitation frequency; the improvement of isolation with increasing frequency in essence is canceled by the increase in the excitation force.

Inertia Bases Vibration-producing machines are often mounted on massive supports, called "inertia bases," to increase the isolated mass. If the total stiffness of the isolators that support the machine is not changed as the isolated mass is increased, then the natural frequency f_n of the system is reduced and the force transmissibility at high frequencies is reduced accordingly. However, limitations on the loads that practical isolators can support often make it necessary to employ more and/or stiffer isolators if an inertia base is added, thus limiting the natural frequency changes and the transmissibility reductions achievable by the addition of inertia bases. However, the increased isolated mass still serves to reduce the vibrational excursions of the machine produced by the forces that act on it.

Limitations of Linear Spring–Mass Model with Prescribed Excitation The classical linear spring–mass model of Fig. 1 provides some insights but does not account for all practical physical phenomena. Real isolators are not massless, may exhibit internal resonances, and may be nonlinear. Real machine frames and supporting structures are not rigid; they may deflect, exhibit multimodal responses and resonances, and rock or otherwise move along paths that are more complex than simple straight lines. In classical analyses, the magnitude of the excitation is taken as independent of the resulting motion, whereas in real vibration sources the motions and forces they produce are interdependent. Some of the effects of these complications are discussed in the following sections of this chapter.

2.2 Isolation of Three-Dimensional Mass

In the simple model of Fig. 1 the mass is constained to move vertically without rotation, the spring–mass system has a single natural frequency, and good isolation is obtained if this natural frequency is much lower than the excitation frequency. In contrast, a three-dimensional mass can translate and rotate with respect to three axes; such a spring-supported rigid mass has six natural frequencies, and good isolation is obtained only if all six of these fall considerably below the excitation frequency. Means for calculating all of the natural frequencies and the vibration responses of general rigid masses are available in Ref. 3; the design of isolation systems is addressed in Refs. 4–6.

2.3 General Linear Uniaxial Isolation[7]

Results obtained from the simple model may suffice for many cases where only relatively low frequencies are of interest, but for higher frequencies complications must be taken into account that result from the lack of rigidity of supports and machine frames, from mass and wave effects in isolators, and from loading effects

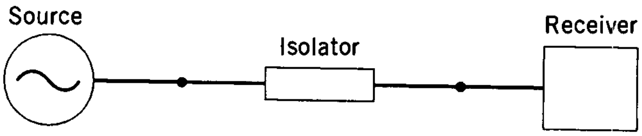


Fig. 2 Schematic representation of general uniaxial isolation arrangement.

on sources. The general three-dimensional situation is extremely complex; fortunately, many practical problems can be treated by focusing on one source-to-receiver connection point and one vibration direction at a time and by considering all components to be mathematically linear. The corresponding general uniaxial isolation problem may be visualized in terms of the diagram of Fig. 2, which shows a vibration source connected to a receiver via an isolator, whose purpose is to limit the magnitude of the vibrations that reach the receiver.

Source Characterization Loading of a source refers to reduction of the source's motion due to forces that oppose this motion. The loading susceptibility of a source may be evaluated by causing the source to act on structures with different dynamic characteristics (impedances) and measuring the velocities and forces produced by the source. For a mathematically linear source, the source's output velocity V_o is related to the force F_o it produces as

$$V_o = V_{\text{free}} - M_s F_o, \qquad M_s = \frac{V_{\text{free}}}{F_{\text{blocked}}}, \tag{4}$$

where all symbols represent complex frequency-dependent quantities. The term V_{free} denotes the source's velocity in the absence of any force generated by the source; F_{blocked} denotes the force produced by the source if its motion is fully blocked (so that the velocity vanishes). The *source mobility* M_s is zero for a *velocity source*, that is, one that produces the same velocity, regardless of its ouput force F_o, and is infinite for a *force source*, that is, one that produces the same force for all output velocities V_o.

General Linear Receiver A general linear receiver may be characterized by its frequency-dependent mobility M_R, which is defined as the ratio of the complex velocity (phasor) at the driving point to the complex force acting at that point. Here, $M_R = 1/Z_R$, where Z_R denotes the receiver's driving point impedance.

Isolation Effectiveness

Definition. Because transmissibility is defined in terms of prescribed excitations, it cannot take account of loading of the source. A measure of isolation performance that is useful in the presence of loading effects is the dimensionless *isolation effectiveness*, which, for a linear receiver, obeys

$$E = \frac{V_{Rr}}{V_{Ri}} = \frac{F_{Rr}}{F_{Ri}}. \tag{5}$$

Here V_R and F_R denote the amplitudes, respectively, of the velocity of the receiver and of the force acting on the receiver, with subscript i referring to the situation where the receiver is connected to the source via an isolator and subscript r referring to the condition where the isolator is replaced by a rigid and massless connection. For *velocity* sources, $E = 1/T$, where the transmissibility T is given by Eq. (1).

Massless Linear Isolator. An isolator may be considered as massless at frequencies that are considerably lower than the isolator's first internal resonance frequency; at such frequencies, the isolator transmits whatever force is applied to it. An isolator is linear if its deformation in the direction of the applied force is proportional to that force. A massless linear isolator may be characterized by its mobility M_I, defined as the ratio of the velocity difference across the isolator to the force applied to it. The effectiveness of such an isolator is given by

$$E = \left| 1 + \frac{M_I}{M_S + M_R} \right| \tag{6}$$

and thus depends on the mobilities of the source and receiver.

For a force source, for which M_S is infinite, $E = 1$. This result indicates that use of a massless isolator in the presence of a force source results in no less receiver motion than use of a rigid connection; in the presence of a more resilient connection the force source generates a greater velocity, keeping the transmitted force unchanged.

Generalized Linear Isolator. An isolator with finite mass cannot be characterized by a single mobility, because it does not simply transmit a force applied to one of its ends. The effectiveness of such an isolator is given by[7]

$$E = \left| \frac{\alpha}{M_S + M_R} \right| \left| 1 + \frac{M_S}{M_{Isb}} + \frac{M_R}{M_{Irb}} \left(1 + \frac{M_S}{M_{Isf}} \right) \right|, \tag{7}$$

where

$$\frac{1}{\alpha^2} = \frac{1}{M_{Irb}}\left(\frac{1}{M_{Isb}} - \frac{1}{M_{Isf}}\right). \qquad (7a)$$

Here, M_{Isb} denotes the isolator's driving point mobility measured on its source side with its receiver side blocked (i.e., constrained not to move), M_{Irb} denotes the isolator's mobility measured on its receiver side with its source side blocked, and M_{Isf} denotes the isolator's mobility measured on its source side with its receiver side free to move.

2.4 Two-Stage Isolation

Whereas conventional single-stage isolation employs a single resilient element between source and receiver, two-stage isolation employs two resilient elements with a mass between them, as shown in Fig. 3. The purpose of the included mass in essence is to provide an inertia force that opposes the imposed vibratory motion, producing better isolation at high frequencies. However, addition of the included mass results in an additional low-frequency resonance at which the isolation performance is degraded.

Mass Isolated by Two-Stage System with Massless Springs

Natural Frequencies. The dynamic system of Fig. 3 has two natural frequencies f_b, given by

$$f_b = f_0(U \pm \sqrt{U^2 - B^2})^{1/2}, \qquad (8)$$

where

$$U = \frac{1}{2}\left(B^2 + 1 + \frac{k_2}{k_1}\right), \qquad B = \frac{f_I}{f_0}, \qquad (8a)$$

$$2\pi f_I = \sqrt{\frac{k_1 + k_2}{m_I}}, \qquad 2\pi f_o = \left[m\left(\frac{1}{k_1} + \frac{1}{k_2}\right)\right]^{-1/2}, \qquad (8b)$$

with k_1 and k_2 representing the stiffnesses of the two isolator springs, m_I the included mass, and m the mass to be isolated; B and U are auxiliary dimensionless quantities. The frequency f_o is the natural frequency that would be obtained if m_I were removed; f_I is the natural frequency that would be obtained if m were fixed. The system's natural frequency f_b obtained with the plus sign in Eq. (8) is greater than both f_o and f_I; the natural frequency f_b obtained with the minus sign in Eq. (8) is less than both f_o and f_I.

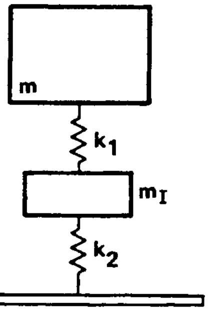


Fig. 3 Mass m and support S isolated from each other by two-stage system consisting of mass m_I between two ideal springs.

Transmissibility. For the system of Fig. 3, the force and motion transmissibilities are equal and obey

$$\frac{1}{T} = \frac{1}{B^2}\left(\frac{f}{f_o}\right)^4 - \left(1 + \frac{1 + k_2/k_1}{B^2}\right)\left(\frac{f}{f_o}\right)^2 + 1. \qquad (9)$$

For excitation frequencies that are considerably greater than both f_o and f_I, the above reduces to $T \approx (f_o f_I/f^2)^2$. At high excitation frequencies, the transmissibility of a two-stage system varies inversely as the fourth power of the driving frequency, whereas that of a single-stage system varies as only the second power. The transmissibility becomes very large (infinite, for an undamped system) if the excitation frequency matches either of the two natural frequencies f_b. For most practical situations, where the damping is not very great, the effects of damping may be neglected, as they are in Eq. (9), if this frequency matching does not occur.

General Uniaxial System Isolated by Two-Stage System with Massless Isolators

Effectiveness. Figure 4 is a schematic sketch showing a general linear receiver with mobility M_R and a general linear source with mobility M_S connected by an isolation system consisting of two massless isolation elements with a mass between them. If M_{IS} and M_{IR} denote the mobilities of the isolation elements located on the source and the receiver sides of the included mass, respectively, and if M_m denotes the mobility of the included mass, the effectiveness of this two-stage isolation system may be written as

$$E = \left|1 + \frac{M_{IS} + M_{IR}}{M_S + M_R} + \frac{(M_{IS} + M_S)(M_{IR} + M_R)}{M_m(M_S + M_R)}\right|. \qquad (10)$$

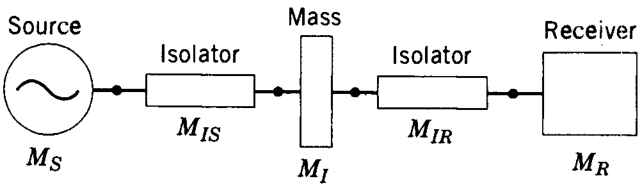


Fig. 4 Schematic representation of general uniaxial two-stage isolation arrangement.

The last fraction in this equation represents the effectiveness increase resulting from inclusion of the mass m_I.

Optimization. If p represents the fraction of the total mobility $M_I = M_{IS} + M_{IR}$ of the isolation elements that is present on the source side, so that the isolator on the source side has mobility $M_{IS} = pM_I$ and that on the receiver side has mobility $M_{IR} = (1 - p)M_I$, then the greatest effectiveness results for

$$p \approx \frac{1}{2}\left(1 + \frac{M_R - M_S}{M_I}\right). \tag{11}$$

In many practical cases the mobility of each of the isolation elements is much greater than the source and receiver mobility; then the greatest effectiveness is obtained for $p = \frac{1}{2}$, for which

$$E \approx \left|1 + \frac{M_I}{M_S + M_R}\left(1 + \frac{M_I}{4M_m}\right)\right|. \tag{12}$$

3 DAMPING OF STRUCTURES

3.1 Effects, Measures, and Measurement

Dynamic System Responses Affected by Damping *Damping* refers to the removal of mechanical energy from a vibration of concern. Although damping has some effect on virtually all responses of structures and other dynamic systems, it has predominant effects on those system motions that are controlled by energy balances rather than by force balances. The system motions that are controlled by energy balances include those associated with excitation at resonance, broadband excitation where the excitation spectrum encompasses one or more system natural frequencies, trace matching (spatial coincidence of exciting pressures with structural surface velocities associated with propagating waves*), freely decaying vibrations, freely propagating waves, and self-excited vibrations (i.e., where the oscillatory exciting force is produced or controlled by the motion).

Measures of Damping

Viscous Damping. In the simple linear mass–spring–dashpot system of Fig. 1, which can also be taken to represent a mode of a vibrating structure, the dashpot represents an element that produces a retarding force that is proportional to the velocity of the mass. The ratio of this force to the velocity is called the *viscous damping coefficient b*. For a system with mass m and spring constant k, the *critical damping coefficient* b_c is given by $b_c = 2\sqrt{km}$; b_c is the smallest value of b for which the aforementioned simple linear system oscillates when it is deflected and released. (If such a system with $b > b_c$ is deflected and released, it drifts toard its equilibrium position without oscillating.)^2 The *fraction of critical damping*, or *damping ratio* ζ, is defined as $\zeta = b/b_c$.

Decaying Free Vibrations. An "underdamped" system (i.e., one with a damping ratio ζ of less than unity) that is disturbed from equilibrium and then permitted to vibrate freely will oscillate with decreasing amplitude, as sketched in Fig. 5. If X_i denotes a peak excursion from equilibrium and X_{i+N} denotes a peak excursion N cycles later, then the *logarithmic decrement* δ is defined as $\delta = (1/N)\ln(X_i/X_{i+N})$.

The rate at which the vibration (displacement, velocity, or acceleration) level decreases with time is called the *decay rate* Δ and usually is expressed in decibels per second. The *reverberation time* T_{60} is defined as the time it takes for the vibration level to decrease by 60 dB, corresponding to a decrease in amplitude by a factor of 1000.

Steady Forced Vibrations. If X denotes the amplitude of the simple system of Fig. 1 produced by a given steady sinusoidal force, then the corresponding amplification A

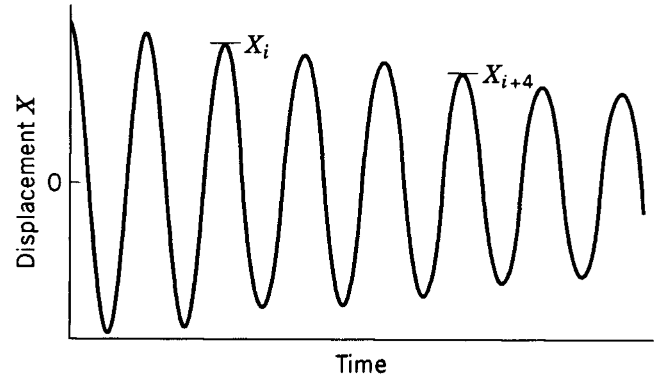


Fig. 5 Time variation of displacement of mass–spring–dashpot system in decaying free vibration. Excursions X_i and $X_i + 4$ illustrate determination of logarithmic decrement.

---

*This occurs when the *wave impedance* Z_t defined in Chapter 64 vanishes.

is defined as the dimensionless ratio $A = X/X_{\mathrm{st}}$, where X_{st} denotes the amplitude that would be obtained if the same force were applied quasi-statically. The *amplification at resonance*, usually called "the Q" of the system and designated by Q, is the value of the amplification that results at resonance, that is, if the excitation frequency matches the system's natural frequency.

The bandwidth of a resonance is defined as the difference between the two excitation frequencies (one above and one below the natural frequency) at which the amplification is equal to $Q/\sqrt{2}$, or, equivalently, at which the amplification is 3 dB lower than Q. The *relative bandwidth* B_f is defined as the ratio of the bandwidth to the natural frequency.

The *phase angle* ϕ by which the displacement of the mass of Fig. 1 lags the excitation force may also be used to characterize the system's damping. For example, at radian frequencies ω much below the system's natural frequency, $\tan\phi = \omega b/k = k_i/k$. Here k_i is the imaginary part of a complex stiffness $k + ik_i$; this complex stiffness, which is equal to the ratio of the phasor of the applied force to that of the resulting displacement, characterizes both the energy storage and the energy dissipation of the system.

Energy Dissipation. The *damping capacity* Ψ of a vibrating system is defined as $\Psi = D/W$, where D represents the energy that is removed from the system per cycle and W denotes the vibrational energy stored in the system. The *loss factor* η, defined by $\eta = \Psi/2\pi$, represents the ratio of the energy removed per *radian* to the stored vibrational energy. In most practical situations the loss factor is less than 0.2 and one may take the stored energy to be equal to the total vibrational (kinetic plus potential energy).

Interrelations. The various measures of damping are related to each other by

$$\eta = \frac{\Psi}{2\pi} = \tan\phi = \frac{k_i}{k} \approx \frac{2.20}{f_n T_{60}} \approx \frac{\Delta}{27.3 f_n} \approx \frac{\delta}{\pi} \quad \text{(in general)}, \tag{13a}$$

$$\eta = 2\zeta = B_f = \frac{1}{Q} \quad \text{(at resonance)}, \tag{13b}$$

where f_n denotes the natural frequency in hertz and T_{60} is in seconds.

Measurement Approaches

Application of Simple Model. Most approaches to measuring the damping of a structure are based on considering a structural mode as a simple mass–spring–dashpot system and applying the foregoing definitions to corresponding parameters measured in freely decaying vibrations or steadily excited sinusoidal motions.

For viscous or approximately viscous damping, the magnitude of the damping present in a system may also be deduced from the phase angle between a sinusoidal force and the resulting motion. This generally is done most conveniently by the use of *Nyquist plots*—plots of the real versus imaginary parts of the system responses at a number of frequencies near resonance[8]—and by means of available software.

Decay Rate in Frequency Band. In applying any approach based on a single-degree-of-freedom model to a structure, care must be taken that the measurement reflects the response of only a single structural mode, with no significant contributions from adjacent modes. An approach that permits measurement of damping in the presence of several modes whose damping does not differ too greatly consists of measuring the rate of decay of a bandpass filtered signal obtained from a vibration sensor that is attached to a freely vibrating structure. (The filter's response must be fast enough so that it can follow the decaying signal. Filters with wider passbands must be used to observe the more rapid signal decays corresponding to higher damping.) For such a measurement it is useful to observe the logarithm of the rectified signal; for a simple viscously damped system, the upper bound envelope of this logarithm is a straight line from whose slope one may determine the damping. Deviations of this envelope from a straight line indicate the presence of several modes with different amounts of damping and/or the presence of other than viscous damping. Most modal analysis and spectrum analysis software includes means for assessing damping from vibration decay measurements based on essentially the foregoing concept.

Steady-State Energy Dissipation Rate. If one applies an oscillatory force at a given frequency f to a structure via an impedance head, one may use the product of the force and velocity signals derived from the impedance head to determine the time-average rate of energy input to the structure. In the steady state, this input rate is equal to the rate of energy loss, which is equal to fD, where D denotes the energy loss per cycle. One may calculate the structure's peak kinetic energy, which is equal to its total vibrational energy W, from measured velocities of the structure and its mass distribution. One may then compute the damping capacity or loss factor by applying their definitions.[9]

3.2 Damping Mechanisms and Magnitudes

Energy Dissipation *Material damping*, also called *mechanical hysteresis* or *internal friction*, refers to the conversion of mechanical energy into heat that occurs in materials as they are deformed. A variety of phenomena may be involved on the molecular, crystal lattice, and metal grain level. Figure 6 indicates the ranges of the loss factors reported for some common materials at small amplitudes and at various frequencies and temperatures.

A vibrating structure may also experience damping as the result of *friction* associated with relative motion between it and structures or fluids in contact with it. An electrically conductive structure moving in a magnetic field is subject to damping due to *eddy currents* generated in it. *Granular materials* in contact with a vibrating structure produce damping as the result of mechanical hysteresis, predominantly at grain asperities, and also as the result of impacts of grains against the structure and against each other. *Impact dampers*, consisting of elements that can rattle against a vibrating structure, essentially extract energy from the structure by converting the structure's vibrational energy to vibrational energy at higher frequencies.

Energy Transport Energy that is transported from a vibrating structure to adjacent structures or fluids (including energy transported in the form of radiated

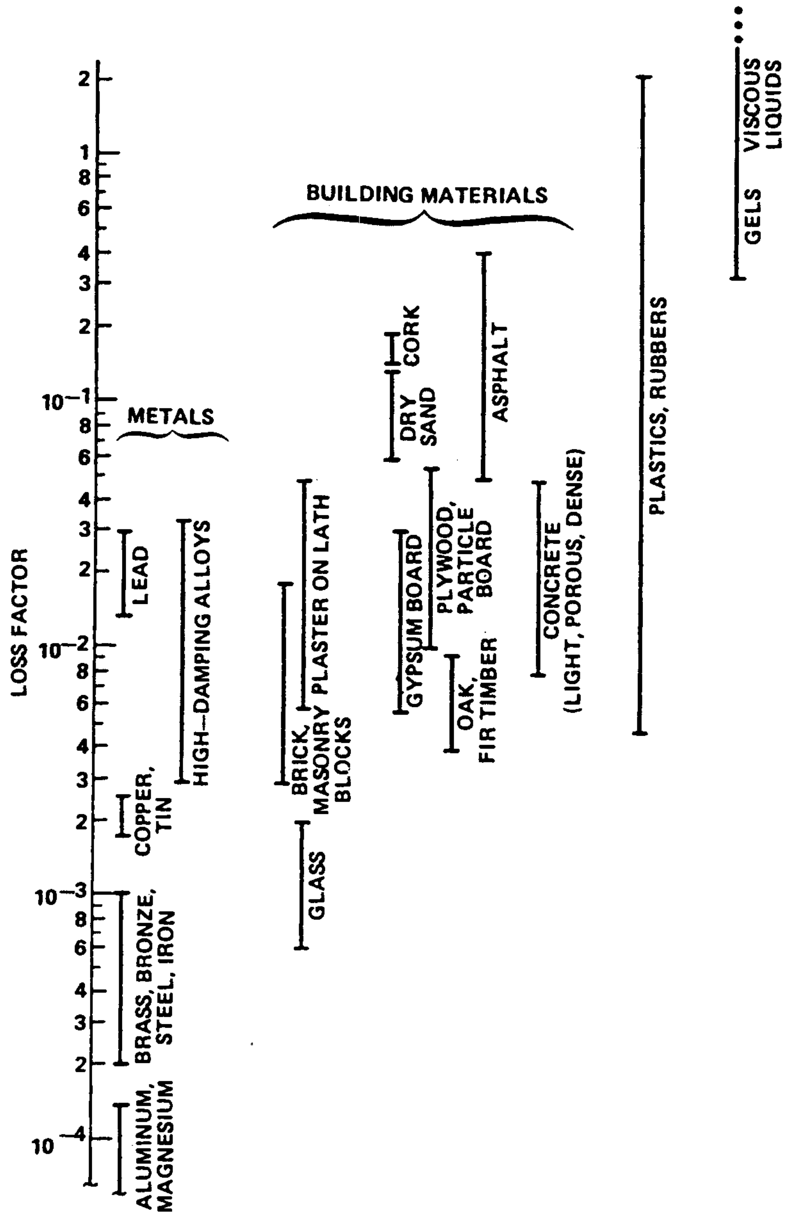


Fig. 6 Ranges of loss factors of some common materials (from Ref. 10).

sound) constitutes a loss of the structure's vibrational energy and thus contributes damping. For a structural component that is in intimate contact with others (e.g., for a panel that is part of an aircraft fuselage), energy transport to adjacent components can contribute considerable damping and generally makes it difficult to measure the component's energy dissipation itself. Energy transport via, for example, supports and cables, usually makes it very difficult in practice to measure the energy dissipation of lightly damped structures.

Boundaries and Reinforcements. For a structural component that behaves essentially as a uniform plate, the loss factor η_b due to energy transport across its boundaries and/or due to dissipation at one-dimensional discontinuities (such as stiffeners or seams) may be estimated from[10]

$$\eta_b = \frac{\lambda}{\pi^2 A} \sum \gamma_i L_i. \tag{14}$$

This relation applies at frequencies that are high enough so that the flexural wavelength λ on the plate is considerably shorter than a plate edge. Here, A denotes the plate area (one side) and γ_i represents the absorption coefficient of the ith boundary or discontinuity element; L_i represents the length of the ith element if that element is along a plate edge or twice that length if the element is in the interior of the plate area; the summation extends over all elements.

The absorption coefficients γ_i may be obtained by applying Eq. (14) to data obtained from loss factor measurements on plates with only one type of element attached. For riveted, bolted, or spot-welded seams or stiffeners, γ_i may be estimated by use of Fig. 7. For seams or stiffeners that are joined by means of continuous welding or by a rigid adhesive, the absorption coefficients generally are negligibly small.

Point-Attached Systems. If a system or structure that initially is free of vibration is attached to a vibrating struc-

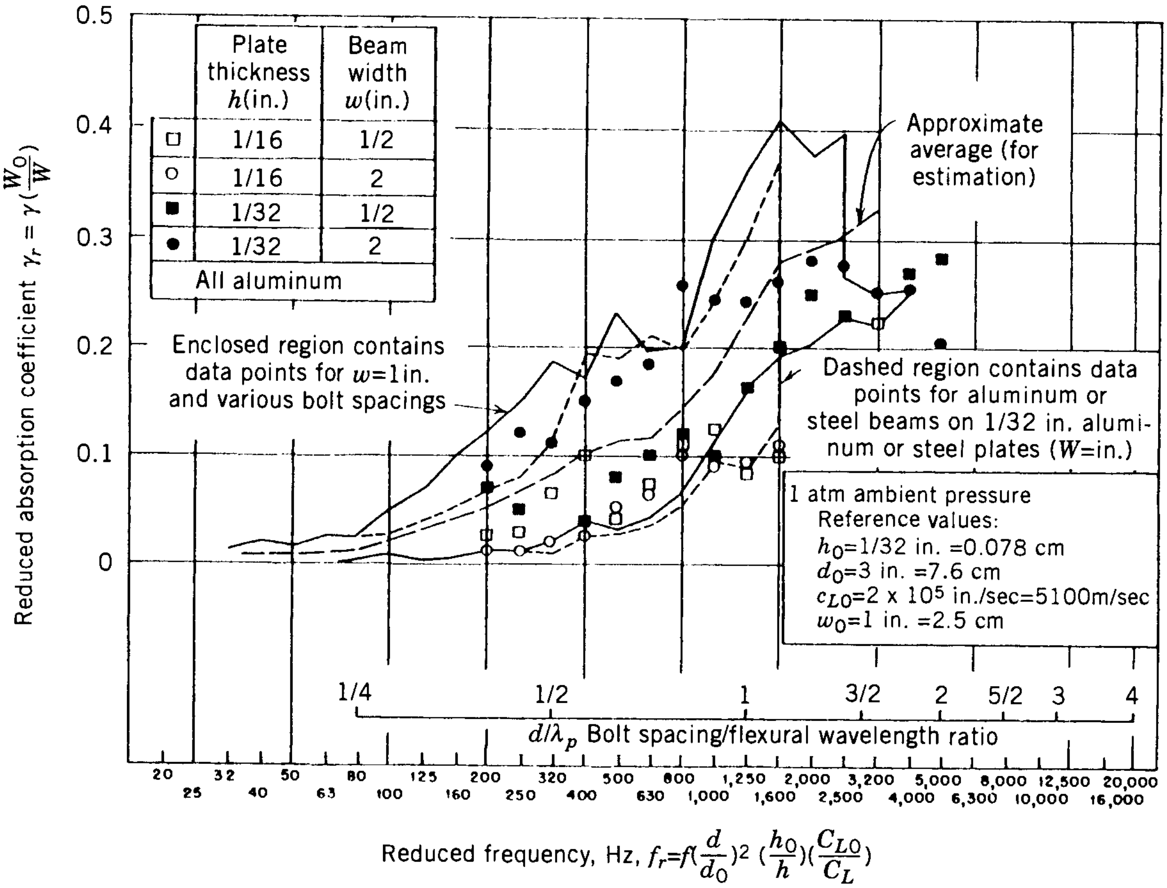


Fig. 7 Reduced data for estimation of absorption coefficients of stiffeners or seams in plates fastened by rivets, bolts, or spotwelds (from Ref. 10): f = frequency, d = fastener spacing, h = plate thickness, w = width of beam in contact with plate or of seam overlap, c_L = longitudinal wave speed in plate material. Subscript 0 refers to reference values.

ture at a point, then the energy D that is transported to the attached structure at frequency f is given by

$$D = \frac{V_s^2}{2f}\,\mathrm{Re}\{Z_A\}\left|1 + \frac{Z_A}{Z_S}\right|^{-2}, \tag{15}$$

where V_s denotes the amplitude of the vibrating structure's velocity at the attachment point before the second structure is attached, Z_A denotes the driving point impedance of the attached structure, and Z_S denotes the impedance of the vibrating structure at the attachment point (with both impedances measured in the direction of V_s).

A *waveguide absorber*[11] essentially consists of a structural element along which waves can travel and which includes means for dissipating the energy transported by these waves. For example, a long slender beam made of a highly damped plastic or coated with such a material may serve as a waveguide absorber at frequencies that are considerably above its fundamental resonance. The damping performance of such an absorber may be evaluated with the aid of the foregoing equation. To be effective, a waveguide absorber must be attached to a structure where it vibrates with considerable amplitude and it must not reduce the structural motion at the attachment point excessively.

Acoustic Radiation. Sound radiated from a structure removes energy from it and thus contributes damping. Acoustic radiation into air rarely results in significant damping, but radiation into liquids may produce considerable damping. The energy D removed from a vibrating structure per cycle at frequency f is given by $D = A\rho c v^2 \sigma / f$, where ρ denotes the density of the ambient fluid, c the speed of sound in it, A the structural surface area in contact with the fluid, v^2 the mean-square normal velocity of the surface area (averaged over both time and space), and σ the radiation efficiency as discussed in Refs. 6 and 12, for example. For a homogeneous plate of thickness h and material density ρ_p, the loss factor corresponding to radiation from one of its sides is given by $\eta = \sigma(\rho/\rho_p)(c/2\pi f h)$. For radiation from both sides, the loss factor is twice the foregoing value.

3.3 Viscoelastic Damping Treatments

Viscoelastic Materials and Material Combinations

Loss Factor of Combinations of Components. A *viscoelastic material* is one that has considerable energy dissipation capability in addition to strain energy storage capability. As evident from Fig. 6, high-strength structural materials generally have little inherent damping, whereas rubbery materials tend to have lesser strength and greater damping. Combinations of structural and viscoelastic materials can take advantage of the strength of one and the damping of the other.

The loss factor η of a structure that is made up of a number of components and that vibrates with a given deformation distribution obeys[10, 13]

$$\eta = \sum \frac{\eta_{ij} W_{ij}}{W_T}, \qquad W_T = \sum W_{ij}, \tag{16}$$

where W_{ij} denotes the strain energy stored in the ith deformation mode (e.g., extension, torsion, shear, flexure) of the jth component and η_{ij} denotes the loss factor associated with that deformation mode and component; the summations extend overall deformation modes and components. The contribution made to the loss factor η of the composite structure by an element deformation is significant only if its loss factor η_{ij} is considerable *and* if its energy storage W_{ij} is a significant fraction of the total energy storage W_T.

Behavior of Viscoelastic Materials. In analogy to the definition of complex stiffness, the complex modulus of elasticity of a material may be defined as $E_c = E + jE_l = E(1 + j\eta_e)$. This frequency-dependent complex modulus is equal to the ratio of the phasor of the applied stress to that of the resulting strain. The real part E is associated with energy storage, the imaginary part E_l with energy dissipation. The material's loss factor η_e associated with the modulus of elasticity is equal to E_l/E. Completely analogous definitions apply for the complex shear modulus.

The viscoelastic materials of greatest practical interest for damping applications are plastics and elastomers. For most practical damping materials of this type, the loss factors associated with all types of deformations are practically equal and the value of the shear modulus is very nearly one-third of the value of the corresponding modulus of elasticity. The various modulus typically vary little with strain amplitude, preload, and aging but vary markedly with frequency and temperature.[13] The moduli typically take on relatively high values at low temperatures and/or high frequencies but take on comparatively small values at high temperatures and/or low frequencies; the greatest loss factors occur in a transition region at intermediate frequencies and temperatures. The modulus values and the transition frequency and temperature ranges vary considerably from material to material. Data on commercially available materials generally may be obtained from their suppliers; a compilation of data for a number of materials appears in Ref. 13.

Beams and Plates with Viscoelastic Additions[10]

Viscoelastic layers typically may be applied to a structure with or without a covering "constraining" structure. In the absence of a constraining structure, the viscoelastic layer stores and dissipates energy primarily in extension/compression; in the presence of a constraining structure the viscoelastic layer acts primarily in shear. A viscoelastic layer system with a constraining structure often can produce greater damping than a viscoelastic layer of the same total weight without a constraining structure, at least in a limited frequency range, but the damping system must be designed for the frequency and structural wavelengths or deflection shapes of concern. On the other hand, a viscoelastic layer without a constraining structure usually can produce reasonable damping performance without taking wavelength into account in the design process.

Extensional Viscoelastic Additions. Figure 8 shows cross-sectional configurations of beams and plate strips where a layer or insert of viscoelastic material is attached to a uniform structural beam or plate so that flexure of the structure causes the viscoelastic component primarily to extend or contract axially. For flexure of such configurations, the loss factor increase η resulting from addition of the viscoelastic materials obeys

$$\eta = \beta\left[1 + \frac{q^2(1+\beta^2) + (r_1/H_{12})^2\alpha}{q[1 + (r_2/H_{12})^2\alpha]}\right]$$

$$\approx \frac{\beta E_2}{E_1}\,\frac{A_2(r_2^2 + H_{12}^2)}{I_2}, \tag{17}$$

where $q = E_2A_2/E_1A_1$; $\alpha = (1+q)^2 + (\beta q)^2$; $r_1 = \sqrt{I_1/A_1}$, $r_2 = \sqrt{I_2/A_2}$; β represents the loss factor of the viscoelastic material; and H_{12} denotes the distance between the neutral fibers of the two components. Here, E denotes the (real) modulus of elasticity, A the cross-sectional area, and I the centroidal moment of inertia; subscript 1 refers to the structural component and 2 to the viscoelastic component. The simpler, approximate form of Eq. (17) applies to the often encountered case where $E_2A_2 \ll E_1A_1$.

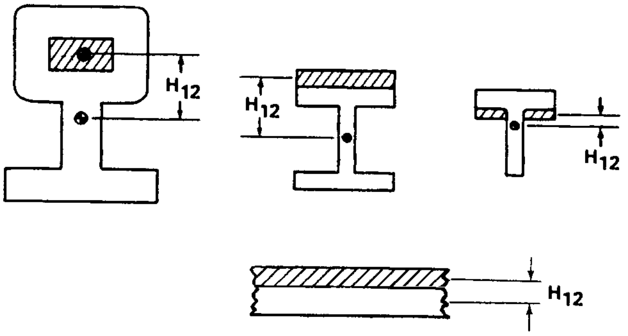


Fig. 8 Cross sections of some beam and plate configurations with attached viscoelastic layers (shown cross-hatched) that extend as the structures flex (adapted from Ref. 10).

Figure 9 shows how the loss factor of a plate with an added viscoelastic layer varies with the layer thickness and loss factor. In general, in order to obtain a relatively large damping increase by adding an extensional viscoelastic element, one needs to use a damping material with a comparatively large βE_2 product in a configuration that produces a significant separation H_{12} between the neutral fibers. In most practical cases the loss factor obtained by adding viscoelastic layers to both sides of a plate is approximately equal to the sum of the loss factors obtained with the layer added separately.

Viscoelastic Additions in Shear. Figure 10 shows generic cross sections of beams and plate strips that include viscoelastic components in configurations for which flexure of the composite causes the viscoelastic element (designated by subscript 2), which is rigidly adhered to two separate structural elements (subscripts 1 and 3), to deform primarily in shear.

The loss factor contribution η resulting from shear of the viscoelastic material may be calculated from

$$\eta = \frac{\beta XY}{1 + (2+Y)X + (1+\beta^2)(1+Y)X^2}, \tag{18}$$

where, for a beam,

$$X = \frac{SG_2b}{p^2H_2}, \qquad Y = \frac{H_{13}^2}{SN}, \qquad S = \frac{1}{E_1A_1} + \frac{1}{E_3A_3},$$

$$N = E_1I_1 + E_3I_3, \tag{19}$$

and for a plate,

$$X = \frac{SG_2}{p^2H_2}, \qquad Y = \frac{12H_{13}^2}{SN}, \qquad S = \frac{1}{E_1H_1} + \frac{1}{E_3H_3},$$

$$N = \frac{1}{12}(E_1H_1^3 + E_3H_3^3). \tag{20}$$

In all cases, β, G_2, and H_2 represent, respectively, the loss factor, shear modulus, and thickness of the viscoelastic material; b represents its average length as measured on a beam cross section; E_i, A_i, I_i, and H_i denote, respectively, the modulus of elasticity, cross-sectional area, centroidal moment of inertia of cross-sectional area, and thickness of component i; and H_{13} denotes the distance between the neutral surfaces of the two *structural* elements indicated by subscripts 1 and 3. Additionally, p represents the wavenumber of the flexural vibration of the composite beam or plate under consideration; it is

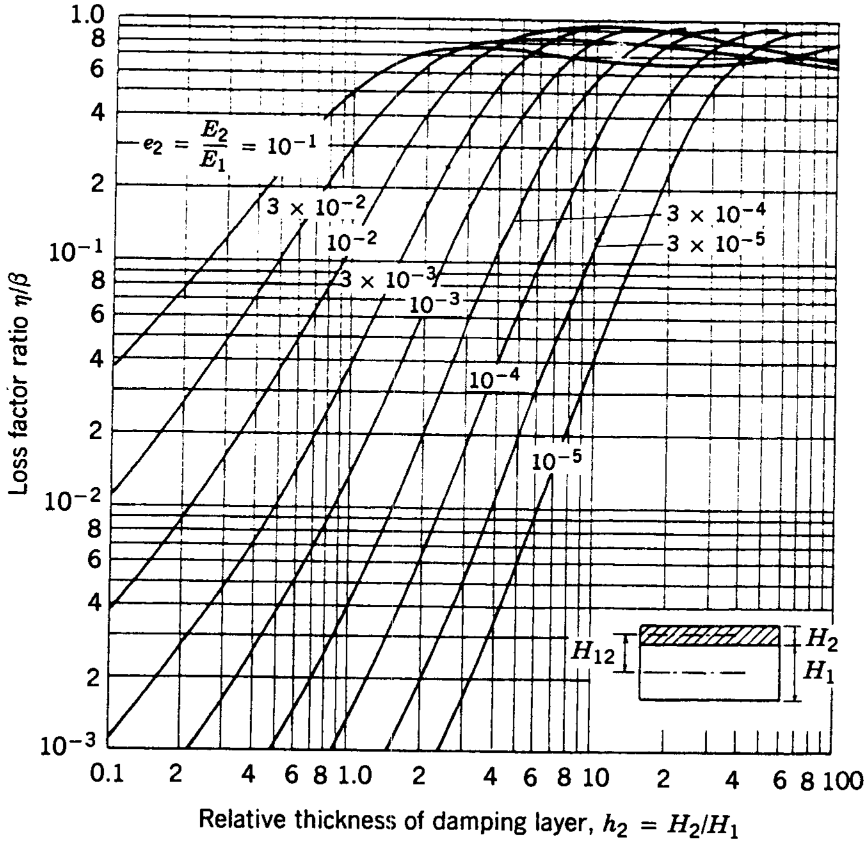


Fig. 9 Dependence of loss factor of plate with extensional viscoelastic layer on relative thickness and relative modulus of layer (from Ref. 10): η = loss factor of plate with attached layer, β = loss factor of viscoelastic material, E = modulus of elasticity, H = thickness. Subscript 1 refers to structural plate, 2 to viscoelastic layer.

related to the wavelength λ at frequency f by

$$\frac{1}{p^2} = \left(\frac{\lambda}{2\pi}\right)^2 = \sqrt{\frac{B/\mu}{2\pi f}}, \qquad (21)$$

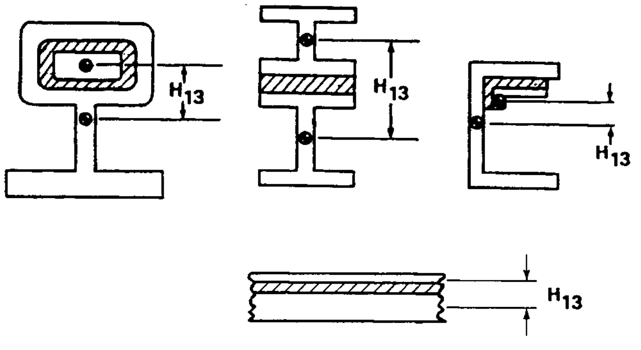


Fig. 10 Cross sections of sandwich plate and of some beam configurations that incorporate viscoelastic elements (cross-hatched) that shear as the structures flex (adapted from Ref. 10).

where B denotes the flexural rigidity of the composite beam (or the flexural rigidity per unit width of the composite plate) and μ represents the mass per unit length of the beam (or the mass per unit area of the plate). The flexural rigidity may be obtained from

$$B = N\left|1 + \frac{YX(1+j\beta)}{1+X(1+j\beta)}\right|, \qquad (22)$$

where $j = \sqrt{-1}$ and the viscoelastic material property values to be used are those that correspond to the temperature and frequency of interest. Where p is not determined by boundary or other constraints, it may be found by solving Eqs. (19) and (20) simultaneously, either explicitly or by a numerical iterative approximation process.[10]

A composite with high damping at a given temperature and frequency may be obtained by use of a configuration with a large value of Y and a viscoelastic material with a high loss factor β, provided the thickness H_2 of that material is selected so that $X = X_{\text{opt}}$, where X is given by Eq. (19) or (20) and $X_{\text{opt}} = [(1+\beta^2)(1+Y)]^{1/2}$. Then the loss factor takes on its greatest value for the applicable Y and β values, namely, $\eta_{\max} = \beta Y/[2 + Y + 2/X_{\text{opt}}]$.

REFERENCES

1. E. E. Ungar, "Equality of Force and Motion Transmissibilities," *J. Acoust. Soc. Am.*, Vol. 90, 1991, pp. 596–597.
2. W. T. Thomson, *Theory of Vibration with Applications*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ, 1981.
3. H. Himelblau, Jr., and S. Rubin, "Vibration of a Resiliently Supported Rigid Body," in C. M. Harris (Ed.), *Shock and Vibration Handbook*, McGraw-Hill, New York, 1988, Chapter 3.
4. J. N. MacDuff and J. R. Curreri, *Vibration Control*, McGraw-Hill, New York, 1958.
5. C. E. Crede and J. E. Ruzicka, "Theory of Vibration Isolation," in C. M. Harris (Ed.), 3rd ed., McGraw-Hill, New York, 1988, Chapter 30.
6. P. M. Norton, *Fundamentals of Noise and Vibration Analysis for Engineers*, Cambridge University Press, Cambridge, 1989.
7. E. E. Ungar and C. W. Dietrich, "High-Frequency Vibration Isolation," *J. Sound Vib.*, Vol. 4, 1966, pp. 224-241.
8. D. J. Ewins, *Modal Testing: Theory and Practice*, Research Studies Press, Letchworth, Hertfordshire, England, 1986.
9. B. L. Clarkson and R. J. Pope, "Experimental Determination of Modal Densities and Loss Factors of Flat Plates and Cylinders," *J. Sound Vib.*, Vol. 77, No. 4, 1981, pp. 535–549.
10. E. E. Ungar, "Structural Damping," I. L. Ver and L. L. Beranek (Eds.), *Noise and Vibration Control Engineering*, Wiley, New York, 1991, Chapter 12.
11. E. E. Ungar and L. G. Kurzweil, "Structural Damping Potential of Waveguide Absorbers," *Transactions, Inter-Noise 84*, Institute for Noise Control Engineering, Poughkeepsie, NY, 1984, pp. 571–574.
12. B. L. Clarkson and K. T. Brown, "Acoustic Radiation Damping," American Society of Mechanical Engineers, Publication 85-DET-24, 1985.
13. A. D. Nashif, D. I. G. Jones, and J. P. Henderson, *Vibration Damping*, Wiley, New York, 1985.

72

VIBRATION MEASUREMENTS AND INSTRUMENTATION

MICHAEL MÖSER

1 INTRODUCTION

Measurements on vibrating mechanical structures cover a broad band of applications ranging from simple laboratory setups to the vibrational characterization of whole bridges, buildings, trains, aircrafts, and so forth. This indicates that vibrational measurements are often complex enough to require a fundamental and detailed understanding of the instrumentation, the sensoring, and the apparatures involved. Normally, measurements are useful only if they are accompanied by some expectation regarding the physical behavior of the tested mechanical system (see all chapters in Part VI and Refs. 1–3).

This chapter will give an overview of the measurement quantities and the different types of sensors. It will address their fundamental principle, their specific use, the limits and inherent problems, and deal with some of the most important applications of vibrational measurements. As detailed overviews Refs. 4 and 5 are recommended for further reading.

2 MEASUREMENT QUANTITIES

The most commonly used measurement quantities are designated the descriptor q for the spatial vibration field at a certain point. These measurement quantities are the displacement $q = \xi(t)$, the velocity $q = v(t)$, the acceleration $q = a(t)$ with the relationships

$$v(t) = \frac{d\xi(t)}{dt} \quad \text{and} \quad a(t) = \frac{dv(t)}{dt}. \qquad (1)$$

The list continues with the angular velocity $q = w(t)$ on the structure, if relevant, the exciting force $q = F(t)$, and the sound pressure $q = p(t)$ in the surrounding medium. The appropriate levels (defined in Chapter 79) are usually based on the root-mean-square (rms) value

$$q_{\text{rms}} = \sqrt{\frac{1}{T}\int_0^T q^2(t)\,dt}. \qquad (2)$$

Sometimes the peak amplitude is used instead of the rms value (e.g., evaluations with Fourier techniques, see Chapter 101). The two values differ by a factor of $\sqrt{2}$. Levels may be expressed by linear or A-weighted filtering methods (Chapter 155), the latter being more convenient for the measurement of emission-levels. The usual time weighting (fast, slow, impulse, see Chapter 155) may also be employed.

Since the kinetic measurement quantities displacement, velocity, and acceleration are related by time differentiations and integrations [(Eq. 1)], their electric representation may be converted into one another by electrical networks or numerically after analog-to-digital conversion (Chapter 101). Integration networks of *RC* type are most commonly used. Their frequency response equals that of an ideal integrator, but only above some cutoff frequency (typically 10 Hz). Therefore a corresponding signal distortion of transients may cause difficulties. Numerical time integration and differentiation can be performed both in the time and frequency domain, the latter by (repeated) division and multiplication with $\pm j\omega$ (note that manufacturers might use different time conventions). In both domains, the mathematical operation causes significant frequency-dependent amplification of the signals noise component.

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker

Copyright © 1997 John Wiley & Sons, Inc.

The above measurement quantities q are the only ones that can be measured directly by using a single (electromechanical) transducer. Indirect methods (based on theoretical assumptions and the use of several transducers) are necessary for intensity and power flow measurements (sound intensity, Chapter 156; structural intensity[6,7]).

3 MEASURING SENSORS

An instrument for the measurement of vibrations consists of one or more electromechanical transducers with a linear relationship between the mechanical input and the electrical output. The electrical path contains a circuit or system to process the transducer's output into a properly defined and correlated indication of the measurement quantity. The sensors should provide the lowest possible sensitivity to environmental effects such as high sound levels (for vibrations), transverse fields, temperature fluctuations, and electromagnetic or other radiation fields.

The dynamic range of the instrument must allow distortion-free measurements of the highest encountered vibration level. The lowest signal-to-noise ratio must exceed 10 dB for an accurate measurement result.

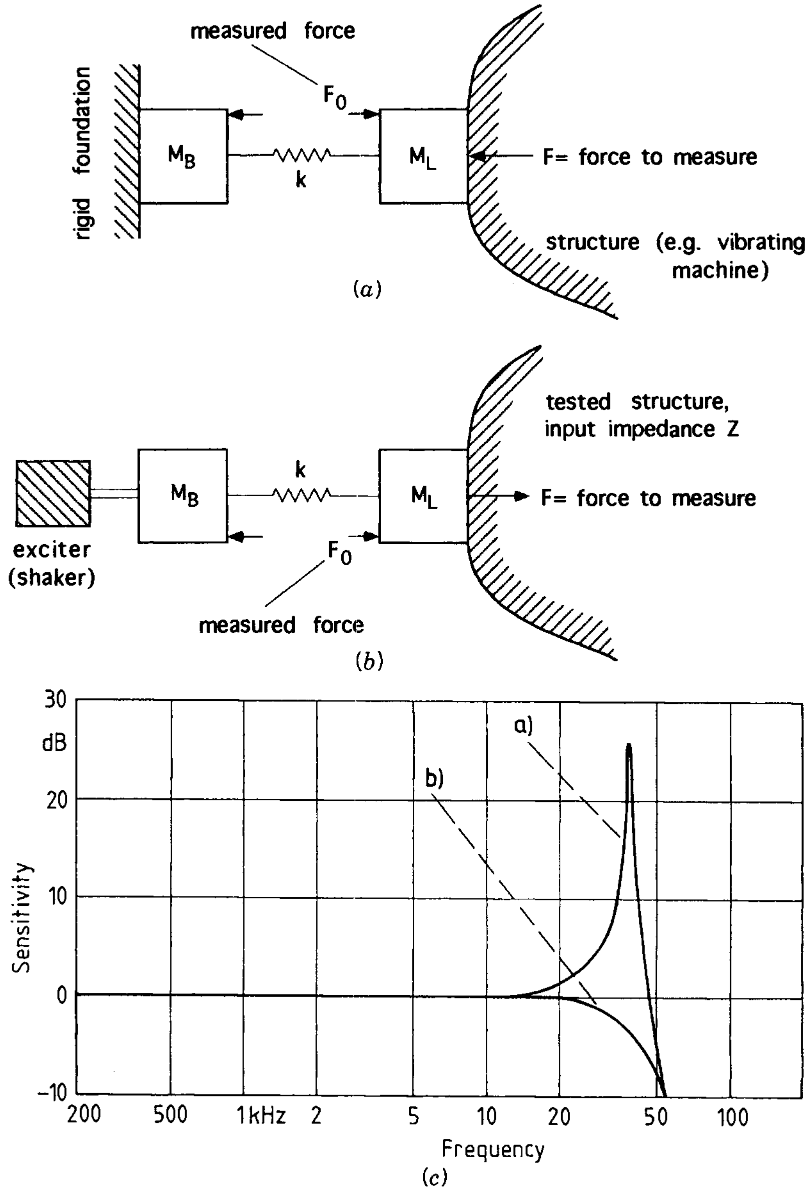

measured force
rigid foundation
F_0
M_B
k
M_L
F= force to measure
structure (e.g. vibrating machine)
(a)
tested structure, input impedance Z
k
M_B
M_L
F= force to measure
exciter (shaker)
F_0
measured force
(b)
30
dB
20
10
0
-10
Sensitivity
a)
b)
200
500
1kHz
2
5
10
20
50
100
Frequency
(c)

Fig. 1 Frequency responses of force transducers: (*a*) Measurement of forces produced by some vibrating structure and transmitted to a foundation, (*b*) measurement of forces exciting some structure, and (*c*) typical frequency responses in case (*a*) and (*b*).

3.1 Piezoelectric Transducers

The most commonly used vibration transducers are of the piezoelectric type. The deformation of a crystal element, consisting of materials such as quartz or artificially polarized ceramics, results in an electrical charge that is proportional to the acting force.[8, 10]

Force transducers are this principle directly. Mechanically, they may be considered to be a spring (the piezoelectric element) with a mass on each end (base M_B and load M_L, which are parts of the housing). Their frequency response depends on the application. For the measurements of force F delivered through the sensor from a structure to a foundation (Fig. 1*a*), the sensor behaves like a simple resonator, giving a measured force $F_0 = F/(1 - \omega^2/\omega_0^2)$, shown in Fig. 1c (upper curve). In this case, the resonant frequency ω_0 depends on the piezoelectric spring with the spring constant k and M_L : $\omega_0^2 = k/M_L$.

For the measurement of forces that are delivered through the sensor to a structure (Fig. 1*b*), the top of the transducer is loaded with an impedance Z, which represents the structure including the sensor's mounting. The measured force F_0 is the difference between the force F acting on the structure and the inertia due to the load M_L :

$$F_0 = F - M_L a_L = F(1 + j\omega M_L/Z), \tag{3}$$

where a_L denotes the acceleration of M_L. Since the mechanical behavior of Z can be associated with a local spring constant k_L at higher frequencies, the measured force in this case is $F_0 = F(1-\omega^2/\omega_0^2)$, where ω_0 denotes the resonant frequency of the local stiffness–load system alone: $\omega_0^2 = k_L/M_L$ (lower curve in Fig. 1*c*). Normally, the relevant mass M_L is minimized in the sensor's construction; foot and head should therefore not be exchanged in the mounting, even if the polarity change is irrelevant.

A piezoelectric accelerometer is made up of a mass (the so-called seismic mass) that is connected to the piezoelectric element (Fig. 2). The piezoelectric element in this case acts as a spring. The force transmitted by the spring due to Newton's law is proportional to the acceleration of the mass. The electrical charge is therefore given by

$$Q = Q_0 \frac{1}{d\omega_0^2} \frac{a}{1 - \omega^2/\omega_0^2}, \tag{4}$$

where a denotes the acceleration of the housing, Q_0 is a constant, $\omega_0 = 2\pi f_0 = \sqrt{k/M}$ is the resonant frequency of the mass–spring system, and k denotes the spring constant of the spring with the thickness d. A typical frequency-response curve is given in Fig. 3.

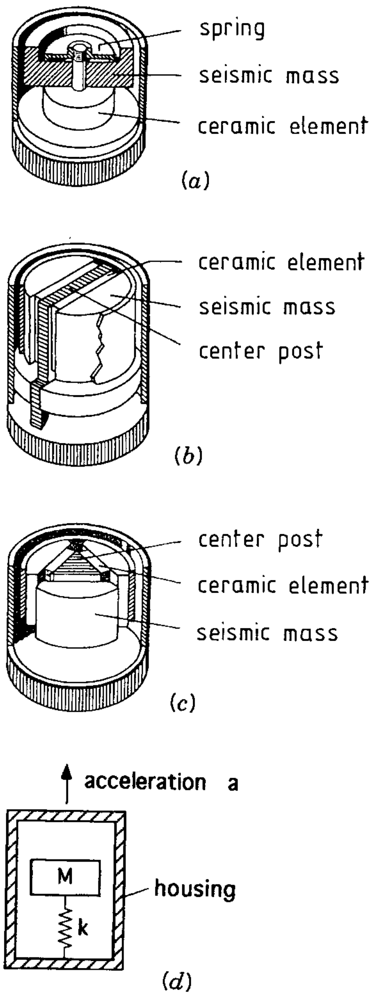


Fig. 2 Types of piezoelectric accelerometers: (*a*) Compression type; (*b*) planar–shear type; (*c*) triangular–shear type; (*c*) triangular–shear type; and (*d*) principle of mechanical behavior, M denotes the seismic mass and k the stiffness of the ceramic element with thickness d.

Below resonance, the foot of the sensor and the mass are strongly coupled. The charge then is directly proportional to the acceleration and to the seismic mass M. This is essentially the reason for the fact that the sensitivity of accelerometers increases linearly with their weight, giving them the disadvantage of lower resonant frequencies. The usable frequency region ends at the point where the frequency response rises to 10% above the flat-line level. Equation (4) gives $f = f_0/3.2$ for that limit.

Transducers with high sensitivity are heavy. They have lower resonant frequencies and a smaller usable frequency range as opposed to light-weight broadband accelerometers with lower sensitivity. Due to the different needs of the various measuring purposes, a range of accelerometers exists, making it possible to choose the correct accelerometer for a given purpose (see Chapter 72). Examples are given in Table 1.

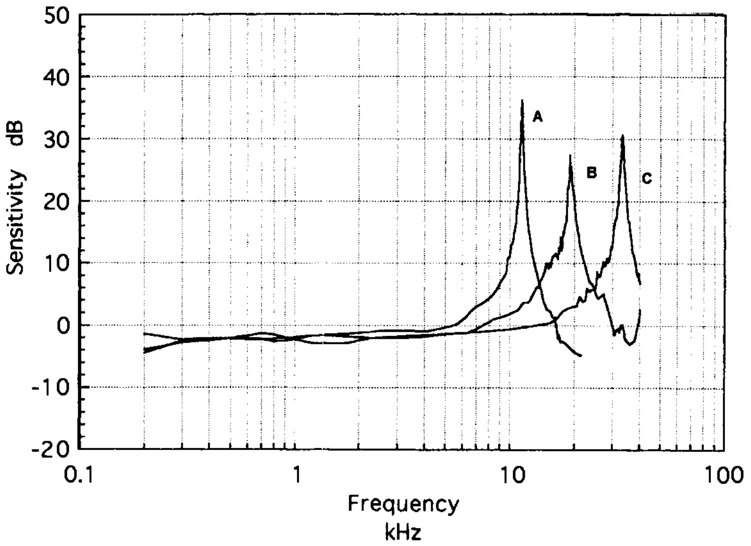


Fig. 3 Typical measured frequency responses of piezoelectric accelerometers, seismic masses: A, 43*g*; B, 30*g*; C, 11.3*g*.

Piezoelectric accelerometers may be designed as compression-type or as shear-type sensors (Fig. 2). The first collect the deformation-induced electrical charge in the material's polarization direction; the second collect the charge perpendicular to the polarization direction.[9] Extra charge due to temperature fluctuations is not picked up with the shear type. This is an important advantage of shear-type sensors, which reduces the temperature sensitivity by a factor on the order of 0.01. Sometimes, the sensitivity to transverse accelerations perpendicular to the sensors main axis causes problems. One example is a composed vibration field with different wave types to be observed. The transverse sensitivity of piezoelectric accelerometers is in the 5% range of the main axis sensitivity. The directivity of the transverse sensitivity depends strongly on the sensors construction (compression, shear, planar shear, and others).

The next element in the measurement system is a charge amplifier of low-input impedance chosen because of the small charge produced by the electromechanical transducer.

Care must be taken in the choice and mounting of the connecting cable to the sensor. The isolation material inside the coaxial cable may locally lose intimate contact to a conductor during cable bending. This, due to the piezoelectric properties of the isolator, may generate a triboelectric charge flow in the conductor across the electrical termination. Cables with noise reduction treatment (a coating technique) prevent triboelectric flow. Furthermore, relative movement between the electrical connec-

TABLE 1 Technical Data of Some Sensors

Type[a]	Sensitivity	Weight (g)	Frequency Limit (kHz)	Dynamic Range (*g*)
PE, S	1.1 pC/g	0.65	26	0.003–5000
PE, DS	3.1 pC/g	2.4	16.5	0.0008–5000
PE, DS	9.8 pC/g	11	12.6	0.0003–5000
PE, DS	31 pC/g	17	8.4	0.00015–2000
PE, DS	98 pC/g	54	4.8	0.00003–2000
PE, DS	310 pC/g	175	3.9	0.000015–500
ICP	10 mV/g	2	10	0.01–500
ICP	100 mV/g	55	3	0.001–50
ED	14 V/ms^{-1}	246	0.2	—
ED	30 V/ms^{-1}	450	2	—

[a] PE, piezoelectric; S, shear type; DS, delta-shear type; ICP, integrated circuit piezoelectric; ED, electrodynamical.

tion and the structure should be avoided by fixing the cable on the tested structure. The cables should leave the object at a nonvibrating point. They should not be longer than necessary because the noise produced by the charge amplifier increases with the cable length.

Integrated circuit piezoelectric (ICP) accelerometers are more comfortable. Their concept is based on the combination of the piezoelectric element and a miniature electrical impedance converter in a common housing with neglectable increase of weight. This combination offers a low-cost, robust measurement channel with fixed sensitivity (independent of cable length), high output voltage (up to 10 V), high signal-to-noise ratio (up to more than 90 dB), and "normal" coaxial cable connections.

3.2 Other Electromechanical Transducers

Corresponding with the various principles for electromechanical transducers used in the measurement of vibrational quantities,[8-10] for example, the variation of inductances or capacitances, there are two more commonly used types commercially available: the geophones, which use the electrodynamic principle of induction, and the strain gauges.

Electrodynamical Sensors Electrodynamic transducers[10] consist of a permanent magnet in form of a hollow cylinder, in the center of which a coil with a large number of turns is suspended by means of light springs (Fig. 4). The output voltage $U \approx \text{const}\ x V_R$ due to electromagnetic induction is proportional to the relative velocity V_R between the coil and the magnet, the latter being fixed in the housing. Since the coil is dynamically decoupled from the magnet above the resonance of the coil-spring oscillatory system, the relative velocity $V_R \approx V$ equals that of the transducers foot V in the current frequency range. Below resonance, the relative velocity tends to zero. This is because of the strong mechanical coupling, which gives a fast decay of sensitivity with decreasing frequency. The resonant frequencies can be achieved at low values, typically between 2 and 20 Hz. To suppress the mechanical resonance, the sensors may be electrically damped by a shunt and by winding the coil on a metal form. This provides damping through induced eddy currents.

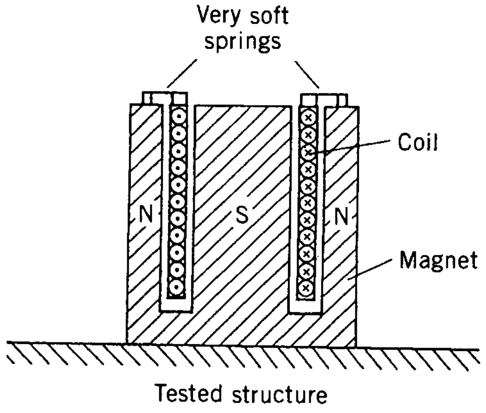


Fig. 4 Construction of electrodynamical sensors.

For the acoustical frequency range, electrodynamic transducers do not have a very constant sensitivity. This is due to their electrical network, which essentially forms a slow decaying low-pass filter. For this reason and due to their weight, electrodynamic transducers serve only for measurements on heavy structures such as buildings, bridges, and for ground vibration measurements. They therefore are frequently referred to as geophones or seismometers.

Strain Gauges This kind of sensor is purely passive. It consists of a metal foil or wire that is fixed on the surface of a test structure. The change of the sensors electrical resistance caused by the strain modulates the output voltage of a suitable electrical network, the gauge being part of it.[10,11] The transducer therefore is nonreversible.

To achieve sensitivities high enough, the gauges consist of many parallel turns (Fig. 5) of foil or wire having small cross sections. The foil or wire can be made of metal or a semiconductor. The axis of main sensitivity is parallel to the lines. Since the changes in resistance are very small, the evaluation of the resistance requires a Wheatstone bridge (Fig. 6), in which the nonactive control resistances serve to equalize the bridge.

The diagonal voltage is proportional to the strain on the gauge. This is represented by R_1 for the so-called quarter bridge with one active branch. The sensitivity can be increased with a half bridge and altering the equalization control resistances. Special use of a series circuit of two transverse gauges with the same cross sections and a length ratio equal to Poisson's ratio of the material under test allows the measurement of tension instead of strain (a result that can also be achieved using two equal gauges forming a specific angle). Measurement of torsional and radial strains are made with special gauge constructions. To compensate for temperature fluctuations, passive gauges may be incorporated in the control resistances. The frequency range of strain measurements is largely limited not by the gauge itself but by the electrical equipment and other considerations, including the vibrational wavelength with respect to the transducer dimensions. Commercially available instruments for use with strain gauges may use a DC or AC voltage to supply the Wheatstone bridge. In the latter case, *in which the* supply voltage is modulated by the strain, the measurement frequency range is limited to 0.1–0.2 times the car-

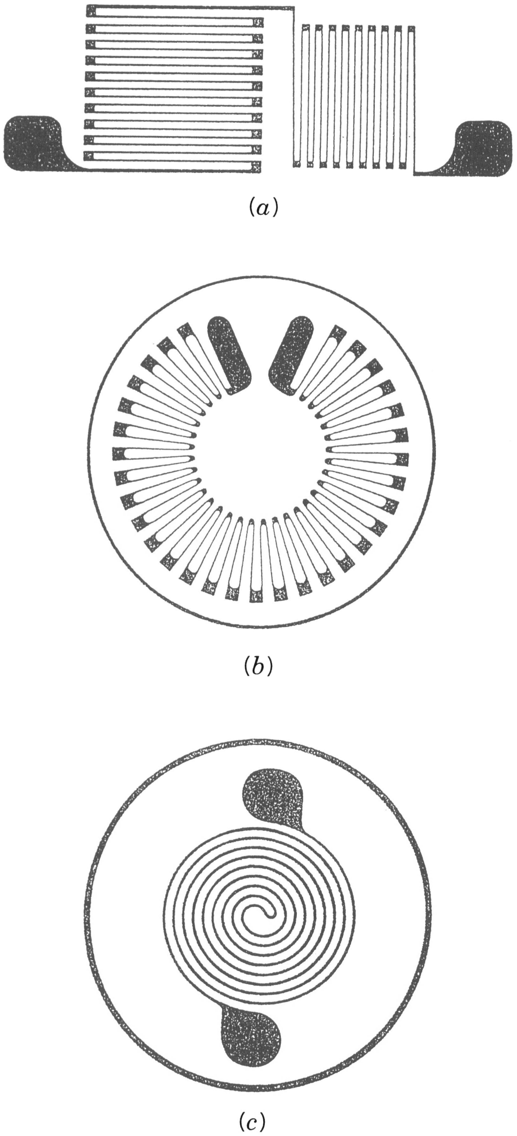


Fig. 5 Examples of strain gauges for the measurement of (*a*) tension, (*b*) radial, and (*c*) torsional strain.

rier frequency. Mainly due to their better stability, carrier systems are preferred. The handling of instruments and mounting of gauges is quite time consuming and requires very careful preparation. Thus, typical applications are long-term observations, for example, at critical points on dynamically loaded shafts and levers or on bridges and other buildings.

Strain gauges made of piezoelectric material are used to provide so-called piezoresistive accelerometers.[10] Their construction principle (similar to that of piezoelectric accelerometers) consists of a seismic mass connected to the housing by an arrangement with a finite spring constant (e.g., by a cantilever beam), carrying the strain gauge. Because of the piezoresistive accelerometer's low sensitivity, no small accelerations can be measured (lower range typically more than $1g$).

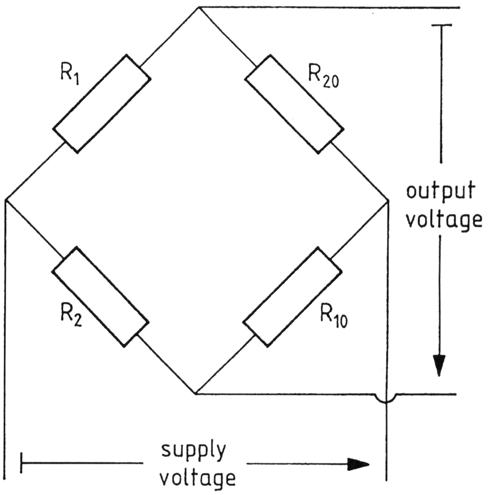


Fig. 6 Wheatstone bridge for strain measurements. R_1, R_{10}, R_2, and R_{20} may represent active or passive resistances (active resistances: quarter-bridge R_1, half-bridge R_1 and R_2, full-bridge: all).

3.3 Choice and Mounting

The correct choice of an accelerometer, or any other vibration sensor, is a question of matching the sensor to the tested structure and ensuring that it does not significantly alter it.

Generally, the mechanical input impedances (moment and force) of the structure at the measured point should be altered as little as possible by the connected sensor. For plates the highest measurement frequency depends on the mass M of the sensor and the impedance of the plate. The highest frequency is given by

$$f < \frac{0.36\rho c_L h^2}{M\sqrt{10^{\Delta L/10} - 1}}, \qquad (5)$$

where ΔL denotes the relative error (dB); ρ, c_L, and h are the density, (longitudinal) wave speed, and thickness of the plate, respectively. If $\Delta L = 1$ dB, then Eq. (5) states that the sensor's mass M should be smaller than the mass of a plate disk with the radius of one-sixth of the bending wavelength. On a metal sheet of thickness of 0.5 mm (as is found in cars), a 2.4-g sensor can be used up to 750 Hz.

Clearly, for measurements on light-weight structures, such as metal sheets, small electromechanical sensors are needed. Heavy structures, like walls, allow the use of larger transducers with higher sensitivity. Often vibration levels of thin structures are quite high and broad

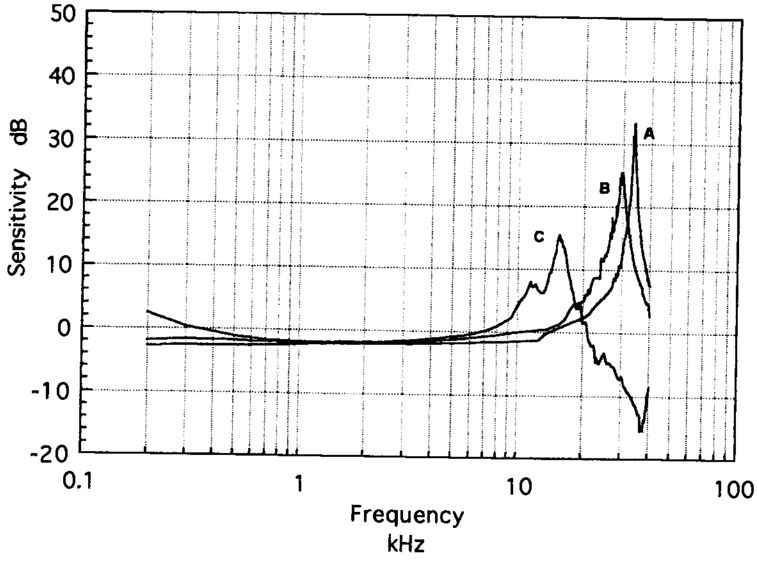


Fig. 7 Influence of sensor (43*g*) mounting, with (A) stud, (B) beeswax, and (C) handheld.

band, thus the need for transducers with high sensitivity occurs mostly for heavy structures, where normally the frequency range of interest is not too high. In most cases the broad range of available sensors (Table 1) allows a correct choice. For special purposes, such as for measurements on membranes or moving parts such as crankshafts, a sensoring is necessary where the structure is not touched (see Chapter 72). A topic is often discussed in the proper attachment of sensors to the structure. Clearly, a screwed connection is preferable. Figure 7 shows that small sensors can be attached using beeswax with almost the same result. In high frequencies are not of interest, double-sided self-adhesive tape or magnetic attachments will suffice. Heavier transducers can be mounted using good (two-component) adhesives. Hand-held monitoring often gives poor, irreproducible results.

3.4 Calibration Techniques

Vibration transducers may be calibrated using one of the following methods:

1. *Use of calibrators.* The sensitivity results from the output signal of the sensor, which is attached to an electromechanical vibration exciter with a previously known single-frequency vibration level. Exciters are available in the form of pocket-sized, battery-powered instruments that can be used where high precision is not required or as exciter tables.
2. *Comparison of the transducer's output* with that of a sensor with known sensitivity, measuring the same vibration. Care must be taken that vibrational differences between the two sensors are excluded. The influence of local mounting, such as stiffness variations, should be avoided. This can be achieved by arranging the transducers "back to back."
3. *Absolute methods* involve the measurement of the sensor's "true" vibration or transmission factor. For example, this could be done using a microscope and observing the amplitude of the (sinewave) displacement. Normally, a Michelson interferometer is used to determine the correct sensitivity. Since the instrumentation needed is so specialized, this kind of calibration is performed only for factory-certified sensitivities or for other high-precision requirements. Another absolute method is the reciprocity calibration[10] (Chapter 18). The standards[12-14] should be noted.

3.5 Optical Methods

Laser Vibrometers Laser vibrometers consist of three essential elements: (1) a radiation unit that produces one (or two) light beam(s) (the width of which is typically 5 μm) to illuminate the measuring point (or points); (2) an optical sensor, whose purpose is the monitoring of the reflected light; and (3) a signal processor, which converts the vibration information, hidden in the back scattered light, into an analog (or digital) output voltage. The basic principles for this conversion are:

1. Doppler effect, which changes the frequency of the reflected light due to the velocity of the observed scatterer

2. The evaluation of the interference stripes in the light pattern by counting the bright–dark transitions

These methods are often referred to as the corresponding codings of frequency modulation (FM) and amplitude modulation (AM), respectively. It should be noted that neither method uses the intensity of the back-scattered light as information. For FM, the output voltage is proportional to the observed velocity (two beams: difference of velocities), for AM to the displacement. For FM, the lowest measurable signal is a question of noise in the electrical path. The typical resolution is on the order of 1 μm/s. For AM, by simply counting the stripes in the pattern, the resolution is half a wavelength of the laser (approx. 350 nm for red light). Therefore a laser vibrometer may not have enough resolution to measure high-frequency vibrations. An enhancement (up to 30 dB in practice) is possible by using the continuous density information in the light pattern. Since the measurement characteristics merely depend on the internal signal processor, signal-to-noise ratios of 80 dB are often realized. For the same reason, the frequency response may be looked at as completely constant in the chosen frequency range. This is only due to the analog–digital converter.

Theoretically, the operation has no lower bound in frequency. In practice, measurements down to 0.1 Hz are possible. For this method calibration is not necessary.

Laser vibrometers have many convenient features and advantages. They provide very easy handling. Precise location of mirrors and other optical equipment is unnecessary (as is the case with holographic methods, see next section). With vibrometers the distances between the optical parts and the observed structure can vary from centimetres to metres. Using glass-fiber connections, the vibrometers allow applications in inaccessible environments and circumstances, including the observation of rotating and hot structure. Since the mechanical characteristics of the object under test are not influenced by the measurement transducer, very thin shells, membranes, needles, and so forth can be observed. Most optical sensors are of such high sensitivity that even coloring of the structure's surface is unnecessary. Note that the relative motion between the instrument and the illuminated point represents the measured quantity. Laser vibrometers therefore are not useful in situations with ambient motion (e.g., on aircrafts, trains, submarines).

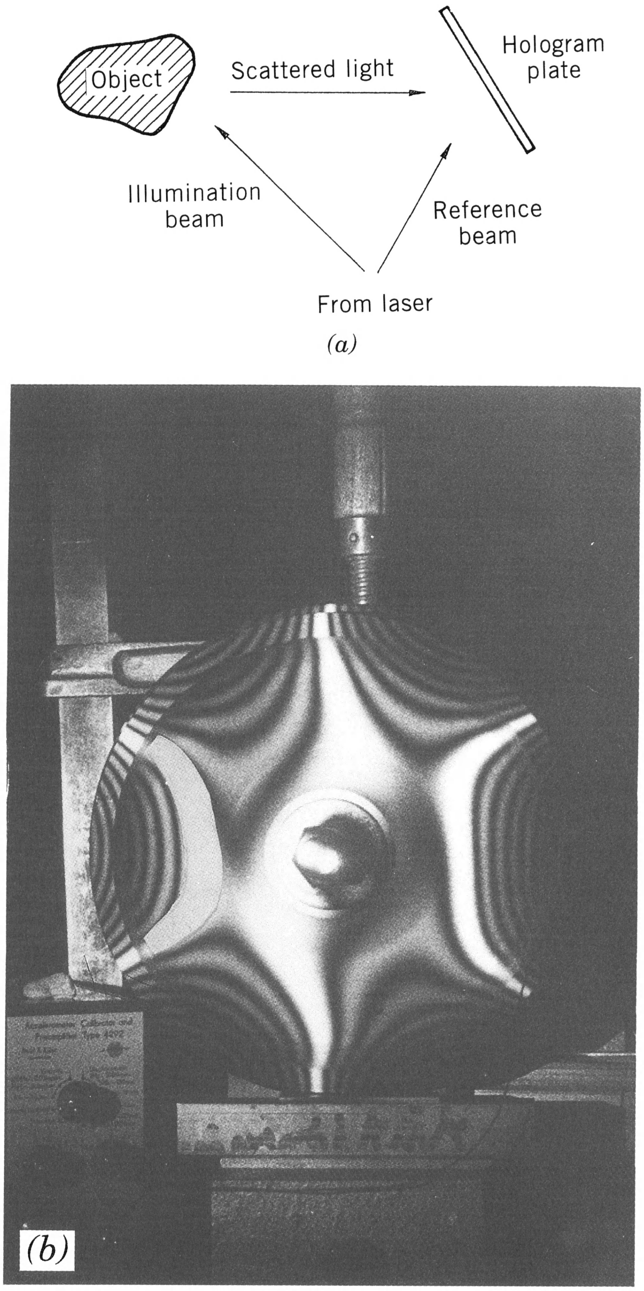


Fig. 8 (*a*) Principle of holography. (*b*) An interference pattern on a vibrating railway wheel model (time-average method used).

Holographic Methods In general, steady-state optical holography (see, e.g., Refs. 15 and 16) is a method that rearranges the scattered laser light field of an undeformed and unmoved body by the illumination of a prefabricated photographic hologram plate (Fig. 8*a*). Therefore, a three-dimensional picture of the body is produced. The hologram plate is exposed before with the originally scattered laser light from the body and a reference laser wave. The processed photographic plate contains the information about amplitude and phase of the originally scattered light, which is necessary to reproduce the spatial light distribution when illuminating the plate with the former reference wave.

Vibrational deformations of a body can be made visible if the hologram plate, for example, is exposed twice in two states of the body's displacements. Then the illumination of the plate simply delivers the sum of the current light distributions, providing an interference pattern with dark and light lines on the picture of the body's surface (see Fig. 8*b*), each line describing a contour of equal displacement. One dark–light–dark transition corresponds to a difference $\Delta\xi$ in the displacement that is proportional to the laser wavelength $\lambda_c (\Delta\xi = \lambda_c/2)$. For periodic, not necessarily single-frequency vibrations, the best resolution is obtained by exposing at the two inversion points. Normally, measurements are made at resonance.

Another method is to expose the hologram plate for a time longer then a cycle of vibration. The reconstructed picture is a result of the time mean of all wavefronts that were scattered from the object during the exposure time. The time of rest in a state of deformation is decisive. This time is longest in the inversion points. The hologram therefore is determined more or less by these two states, the resolution is $\Delta\xi = \lambda_c/4$. A related third method is the exposure of the nonvibrating object and illumination of the hologram plate by both the scattered light from the vibrating body and the reference laser wave. Since the eye (and a photographic picture) integrates only over the absolute intensity, again a pattern of dark and light lines is provided, whose resolution is $\Delta\xi = \lambda_c/2$. The very specialized equipment needed for optical holography makes it an instrument to be used only in special cases. Since the numerical evaluation of the results implies a quite complicated handling, this optical form of modal analysis is normally used to give an illustrated survey of the vibrational shapes.

4 SPECIAL VIBRATION MEASUREMENTS

Frequently, one is confronted with the simple problem of investigating the vibrational characteristics of a structure by measuring suitable quantities. Some examples follow:

- Vibration control of highly sensitive optical or medical equipment
- Surveillance of the vibration of structures in the vicinity of intense construction works (e.g., close to critical pressure vessels or pipelines)
- Investigation of vibration transmission paths
- Determination of acoustical radiation factor

In some cases the forces introduced into a system by an external excitation are of interest but are not directly measureable (e.g., elevators in buildings). This can be determined by an indirect method, using the principle of reciprocity.[1]

4.1 Measurement of Material Parameters

Important mechanical material parameters are those characterizing the specific stiffness (Young's modulus, shear modulus, bending stiffness, etc.) and a quantity describing the internal losses. These parameters are often determined experimentally using resonators, which contain a probe of the material (Table 2). From the resonant frequencies, the stiffness quantities can be computed.

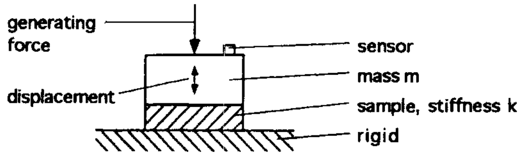
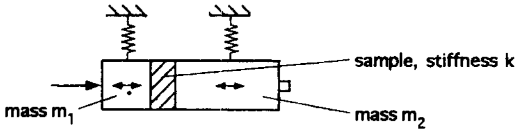
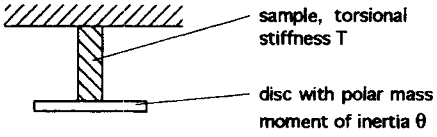
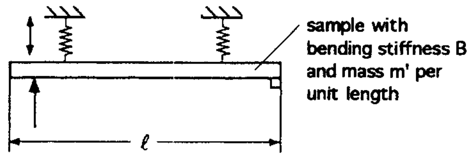
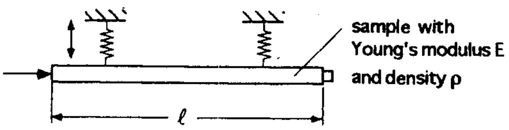
With the same setups, the losses can be found by measuring the half-power bandwidth or, in the case of low loss factors and frequencies, the reverberation time. Other quantities describing losses can be computed by means of Table 3. If the frequency response curves are found with fast Fourier transform (FFT) methods, the half-power bandwidth should contain at least four to six frequency samples.

Only rather small loss factors can be found using resonance experiments, the limit depending on the resonator used. Other methods are available for higher damping. These include:

- Locally sampling the wave field on beams, similar to measurements of absorption coefficients in impedance tubes. This method can be used for weak damping as well.
- Measurement of local level reduction D per metre along beams ($D = 27.2\,\eta/\lambda_L$ for longitudinal waves with length $\lambda_L, D = 13.6\,\eta/\lambda_B$ for bending waves with length λ_B).
- Input impedance measurements (e.g., of the simple mass–spring system).

All these methods are based on a mathematical model for the specific structure with a number of unknown parameters, which are to be determined by the measurements. Generally, the list of mechanical structures, of which the resonators mentioned above are examples, may be enlarged with any structure for which a precise mathematical description is well known. The material parameters then result from one or more fits between measurements and the theoretical model, which, again, is the basic idea of modal analysis (see Chapters 63 and 72). Errors occur due to violations of the assumptions, inherent in the model, or from external influences not taken into account. For example, experiments with the simple resonator must be made under the condition that only the relevant "whole-body mode" is excited and analyzed. For experiments based on specific wave types, transformations into other wave types must be avoided. Losses result not only from internal damping but often from the effects of energy transport over the boundaries or supports. This influence must be reduced or avoided

TABLE 2 Estimation of Mechanical Stiffness from Resonators[a]

Spring–mass system	 generating force displacement sensor mass m sample, stiffness k rigid	$k = 4\pi^2 f_R^2 m$
Mass–spring–mass system	 sample, stiffness k mass m$_1$ mass m$_2$	$k = 4\pi^2 f_R^2 \dfrac{m_1 m_2}{m_1 + m_2}$
Torsional resonator	 sample, torsional stiffness T disc with polar mass moment of inertia θ	$T = 4\pi^2 f_R^2 \theta$
Bending waves in beams	 sample with bending stiffness B and mass m' per unit length ℓ	$B = m' l^4 f_n^2 \dfrac{64}{\pi^2(2n-1)^4}$
Longitudinal waves in beams	 sample with Young's modulus E and density ρ ℓ	$E = \rho l^2 \dfrac{4 f_n^2}{n^2}$

[a] f_R, resonance frequency; f_n, resonance frequency with n nodes; S, cross-sectional area, d = thickness of sample; k, SE/d, E, Young's modulus, ρ = density; l, length.

TABLE 3 Descriptors for Losses

	Loss Factor η	Half-Power Bandwidth Δf	Degree of Damping ϑ	Reverberation Time T	Q factor, Q	Damping Coefficient c	Log. Decrement λ
$\eta =$	η	$\Delta f/f_0$	2ϑ	$\dfrac{2.2}{f_0 T}$	$\dfrac{1}{Q}$	$c/\sqrt{km}$	$\dfrac{2\lambda}{\sqrt{4\pi^2+\lambda^2}}$
Δf	$f_0\eta$	Δf	$2f_0\vartheta$	$\dfrac{2.2}{T}$	$\dfrac{f_0}{Q}$	$\dfrac{c}{2\pi m}$	$\dfrac{2f_0\lambda}{\sqrt{4\pi^2+\lambda^2}}$
$\vartheta =$	$\eta/2$	$\Delta f/2f_0$	ϑ	$\dfrac{1.1}{f_0 T}$	$\dfrac{1}{2Q}$	$\dfrac{c}{2\sqrt{km}}$	$\dfrac{\lambda}{\sqrt{4\pi^2+\lambda^2}}$
$T =$	$\dfrac{2.2}{f_0\eta}$	$\dfrac{2.2}{\Delta f}$	$\dfrac{1.1}{f_0\vartheta}$	T	$\dfrac{2.2Q}{f_0}$	$\dfrac{13.82m}{c}$	$\dfrac{1.1\sqrt{4\pi^2+\lambda^2}}{f_0\lambda}$
$Q =$	$1/\eta$	$f_0/\Delta f$	$\dfrac{1}{2\vartheta}$	$\dfrac{f_0 T}{2.2}$	Q	$\dfrac{\sqrt{km}}{c}$	$\dfrac{\sqrt{4\pi^2+\lambda^2}}{2\lambda^2}$
$c =$	$\eta\sqrt{km}$	$2\pi m\,\Delta f$	$2\vartheta\sqrt{km}$	$\dfrac{13.82m}{T}$	$\dfrac{\sqrt{km}}{Q}$	c	$2\lambda\sqrt{\dfrac{km}{4\pi^2+\lambda^2}}$
$\lambda =$	$\dfrac{\pi\eta}{\sqrt{1-(\eta/2)^2}}$	$\dfrac{\pi\Delta f}{\sqrt{f_0^2-(\Delta f/2)^2}}$	$\dfrac{2\pi\vartheta}{\sqrt{1-\vartheta^2}}$	$\dfrac{6.91}{\sqrt{f_0^2T^2-1.21}}$	$\dfrac{\pi}{Q\sqrt{1-(\frac{1}{2}Q)^2}}$	$\dfrac{2\pi c}{\sqrt{4km-c^2}}$	λ

f_0 = undamped natural frequency = $\sqrt{k/m}/2\pi$, k = spring constant, m = mass.

by the correct choice of supports to decouple the structure under test from the environment. Radiation losses also contribute to the vibration damping. Their loss factor is approximately

$$\eta = \frac{\rho c}{m''\omega}\,\sigma \tag{6}$$

for plates with mass m'' per unit area where σ describes the radiation efficiency, and ρc denotes the specific wave impedance in the surrounding medium. Finally, when interpreting the results, it must be remembered that these results are sometimes based on certain assumptions and are hence not easily transferable to more general cases (e.g., the loss factors of sandwich plates differ significantly for longitudinal and bending waves).

4.2 Modal Analysis

The object of modal analysis is the determination of eigenmode shapes for the object under investigation.[17] This implies that the description of the structure's vibration in terms of modes is meaningful, which is not the case for high damping. No energy should be lost to the surrounding structures and medium, and the internal damping should not be high. Free boundary conditions should be sought when appropriate. This is usually achieved by suspending the test object by very light springs, which results in low-frequency rigid-body modes (modes without elastic deformation of the structure itself). The application of modal analysis ranges from small structures, such as engine crankcases, engine blocks, and other technical components, to large constructions such as an entire car chassis, an aeroplane, or a railway carriage.

In principle, the structure under investigation is set in motion by one or more broadband force inputs, and the resulting motion is observed. One method is to excite the structure in a certain point (e.g., using a shaker) and to measure the structure's response at many points on the surface. Alternatively, the reciprocity principle allows the measurement of the response in one point only with excitation in many points one after another (usually using an impulse hammer). The input and output signals are converted into transfer functions with an FFT analyzer. The transfer functions are then fed into a computer with the modal analysis software. This software is used to calculate resonant frequencies, damping, and the mode shapes by "curve-fitting" methods, for which several different procedures are available (a more detailed description of the theory of modal analysis is given in Chapter 63). Besides the obvious applications of the modal analysis technique (examination of the sound radiation, vibrational discomfort in vehicles, aimed noise control measures, etc.), it also can be utilized to assist in modeling structural modifications. The orthogonality of the modal forms allows for the calculation of the so-called mass and stiffness matrices of the structure. These can be altered by the investigator in a number of different ways and the resulting changes can be seen in the newly calculated parameters and mode shapes.

Modal analysis has a number of inherent problems. They can be caused, for example, by the local properties of the test structure (e.g., local distribution of damping), which cannot be described by a single global value. It also is possible that the use of different curve-fitting methods leads to noticeable differences between modal forms, especially for strongly coupled modes. Therefore, in practice, an error observation analysis is often unavoidable (e.g., the construction of synthetic transfer functions through the predetermined modal parameters and the comparison with the measured curves, test of orthogonality, evaluation of modal parameters and shapes using different curve-fitting methods, etc.).

REFERENCES

1. L. Cremer and M. Heckl, *Structure-Borne Sound: Structural Vibrations and Sound Radiation at Audio Frequencies*, translated and revised by E. E. Ungar, Springer, Berlin, 1988.
2. F. J. Fahy, *Sound and Structural Vibration: Radiation, Transmission and Response*, Academic, London, 1985.
3. M. C. Junger and D. Feit, *Sound, Structures and Their Interaction*, MIT Press, Cambridge, MA, 1986.
4. M. P. Blake and W. S. Mitchell, *Vibration and Acoustic Measurement Handbook*, Spartan Books, New York, 1972.
5. C. M. Harris, *Handbook of Acoustical Measurements and Noise Control*, 3rd ed., McGraw-Hill, New York, 1991.
6. G. Pavic, "Structural Surface Intensity: An Alternative Approach in Vibration Analysis and Diagnosis," *J. Sound Vib.*, Vol. 115, 1987, pp. 405–422.
7. *Proc. 3rd Int. Congress on Intensity Techniques. Structural Intensity and Vibrational Energy Flow*, Centre Technique des Industrie Mécaniques, Senlis, 1990.
8. J. T. Broch, *Mechanical Vibration and Shock Measurement*, 2nd ed., Brüel & Kjaer, Naerum, Denmark, 1988.
9. M. Serridge and T. R. Licht, *Piezoelectric Accelerometers and Vibration Preamplifiers*, Brüel & Kjaer, Naerum, Denmark, 1987.
10. C. M. Harris, *Shock and Vibration Handbook*, 3rd ed., McGraw-Hill, New York, 1988.
11. J. Vaughan, *Strain Measurement*, Brüel & Kjaer, Naerum, Denmark, 1975.
12. American National Standards Institute, "Selection of Calibration and Tests for Electrical Transducers Used for Measuring Shock and Vibration," ANSI S2.11, New York, 1969; rev. 1978.
13. American National Standards Institute, "Methods for the

Calibration of Shock and Vibration Pickups," ANSI S2.2, New York, 1959, rev. 1981.

14. International Standardization Organization, "Methods of Calibration of Vibration and Shock Pickups," ISO 5347, 1985.
15. G. M. Brown, R. M. Grant, and G. W. Stroke, "Theory of Holographic Interferometry," *J. Acoust. Soc. Am.*, Vol. 45, 1969, pp. 1166–1179.
16. J. W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill, New York, 1968.
17. D. J. Ewins, *Modal Testing Theory and Practice*, Wiley, New York, 1984.

73

MACHINERY CONDITION MONITORING

GOVINDAPPA KRISHNAPPA

1 INTRODUCTION

Mechanical condition monitoring is concerned with evaluation of the operating condition of a machinery system or its components and is accomplished through the analysis and interpretation of signals acquired from sensors and transducers. The main purpose of condition monitoring is to detect the presence of faults of damage in machinery during operation. It encompasses both diagnosis and prognosis in order to determine the remaining safe operating life of a machine before a breakdown or failure occurs. *On-condition* operation of machinery leads to efficient maintenance by preventing unexpected failures and improving safety.

Vibration analysis is a reliable technique that is used extensively for condition monitoring by machinery users and operators. Dynamic forces operating within the machinery generate vibratory forces that carry information on the condition of individual components. These vibrations are transmitted via complex paths to the casing, producing external vibration and noise radiation. Sources of vibration include impacts, forces in gears and cams due to machining variations and contact stiffness, electromagnetic forces, combustion pressures, and forces due to air and liquid flow. Detailed discussions on vibrating systems can be found in Chapters 62 and 63. In very simple machines, peak and root-mean-square (rms) values of vibration energy are trended to monitor system condition. Frequency spectra play a valuable role in fault detection, especially in rotating machines, and contain information on operating conditions that reveal malfunctions such as rotor imbalance, shaft misalignment, and bearing and gear damage. In complex machines containing multiple components, more elaborate signal processing and analysis techniques are required to extract information on the component of interest from the overall signal to enable effective detection of early damage.

Machinery diagnostics involve extracting the properties of vibration signals (temporal waveform or frequency spectra) of machinery components operating in good, as well as faulty, conditions. Under both conditions, the properties of signals generated by the individual components are recovered from the overall signature. In addition to energy levels, phase information can be used effectively to diagnose machinery condition in many instances. Statistical descriptors of the wave or spectral forms are useful to identify faults generated by impulsive events.

This chapter describes vibration analysis techniques for condition monitoring of mechanical components. Familiarity with general vibration theory and sensors and instrumentation for acquiring vibration data are required. Some knowledge of the basic tools of signal processing theory and statistical concepts would be helpful. Since this area is the subject of further research and development, only well-established techniques and methods are presented.

2 VIBRATION CHARACTERISTICS OF MECHANICAL FAULTS

Any machinery operation involves the generation *of* forces that produce vibrations. Even a machine in good running condition produces its own characteristic vibrations, caused by various dynamic forces associated with its operation. Sources of various dynamic forces that pro-

---

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker
Copyright © 1997 John Wiley & Sons, Inc.

duce vibrations include: imbalance, reciprocating action, misalignment, impact, sliding, electromagnetic forces, combustion, pressure, fluctuating forces due to air and liquid flows, variations in surface finish of machinery parts, and variations in local contact stiffness.

As machines wear, the play between the parts increases and the individual parts impact much harder on each other, producing stronger vibrations. Bearing wear displaces a shaft from its original axis of rotation to produce imbalance. Worn rotating and sliding surfaces show fatigue cracks. Vibration levels increase due to irregularities in the surface finish of the contacting machine part.

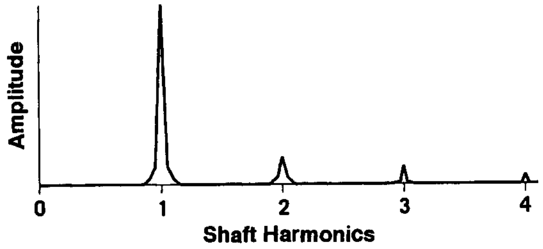
As a machine ages, vibration levels tend to gradually increase with only a slight change in the vibration pattern. Vibrations increase rapidly when machinery components approach the end of their useful life, producing drastic changes in vibration levels and patterns. Although individual faults produce their own characteristic vibration patterns, identification of a fault can be extremely difficult because of vibrations produced by other components in the machine and the presence of complex transmission paths between the component and the measurement point. Some faults on less complex machines can be identified by placing the transducer close to the mechanical component.

2.1 Imbalance

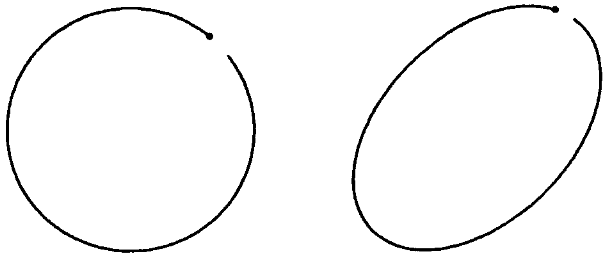
Imbalance is a frequent cause of failure in machines. The most common cause of imbalance is a physical difference between the mass center of gravity and the rotating center of the rotor system.[1] Magnetic forces and temperature gradients also cause shifts in the rotor system that produce imbalance. Imbalance can be characterized by a high-amplitude vibration signal at the shaft rotation frequency, frequently with an elliptical orbit (Fig. 1).[2] Excessive imbalance will also produce an abnormally high peak amplitude at the critical speed of the rotor system, decreasing in amplitude beyond the critical speed. Figure 2 shows that the vibration amplitude will remain higher than that of a balanced rotor at corresponding speeds.

2.2 Misalignment

Shaft misalignment is usually caused by the incorrectly adjusted bearings or couplings. Severe damage could result if care is not taken to correct this fault. Misalignment is expected in all but the most precisely manufactured equipment. Misaligned shaft bearings typically produce vibration characteristics similar to those shown in Fig. 3. Vibration amplitude levels at the first four harmonics of the shaft rotation frequency are high, with a peak at the second harmonic. Misalignment is usually associated with a peak vibration level at twice the shaft rotation frequency. The characteristics of shaft motion for minor, moderate, and severe misalignments are also shown in the figure. Poor alignment can cause bearings to be unloaded to the point where instability is induced. Other problems can include coupling lockup, resulting in thrust bearing failure.



(a) Frequency Spectrum



Symmetric Response Asymmetric Response

(b) Shaft Orbit

Fig. 1 Vibration characteristics of imbalance.

2.3 Mechanical Looseness

Mechanical looseness, the improper fit between component parts, is generally characterized by a long series of high-amplitude harmonics of the shaft frequency (Fig.

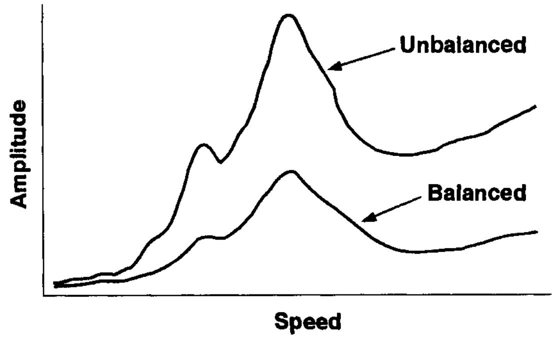
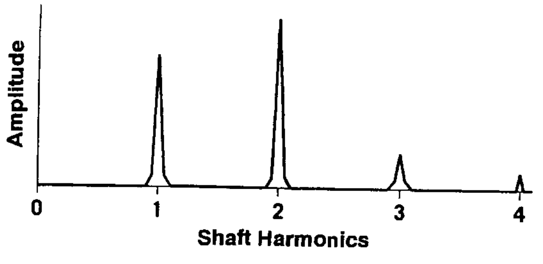
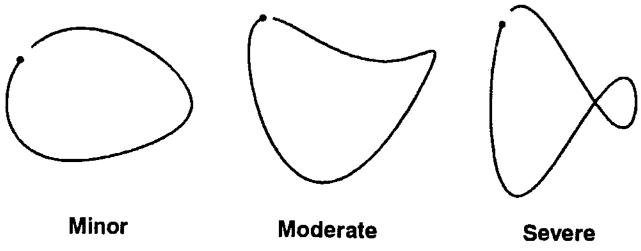


Fig. 2 Vibration amplitude response of balanced and unbalanced rotors.



(a) Frequency Spectrum



(b) Shaft Orbit

Fig. 3 Vibration characteristics of misalignment.

4). Although the exact method by which the harmonics are generated is not well understood, they probably originate from a nonlinear response of the loose part to dynamic inputs from the rotor. Looseness of bearings are also likely to cause vibrations at subharmonics of the shaft rotation frequency (i.e., $f/2, f/3, \ldots, f/n$).[2]

2.4 Rubs

A rub is generally a transitory phenomenon caused by several mechanisms, the most common of which are rotor and seal rubs. Blade rubs induced by insufficient clearance between the blade tips and the casing wall are common in turbomachinery. Bearing wear, caused by problems such as lubrication starvation, imbalance, or vibrations, can also lead to rotor rubs. Rub changes the vibration characteristics of the rotor and its orbit. In the simple case shown in Fig. 5, a flattened waveform with subsynchronous frequency components and a flattened orbit diagram are indications of rub.[2] Because rub is a nonlinear phenomenon, it is not possible to characterize vibration signatures for all cases.

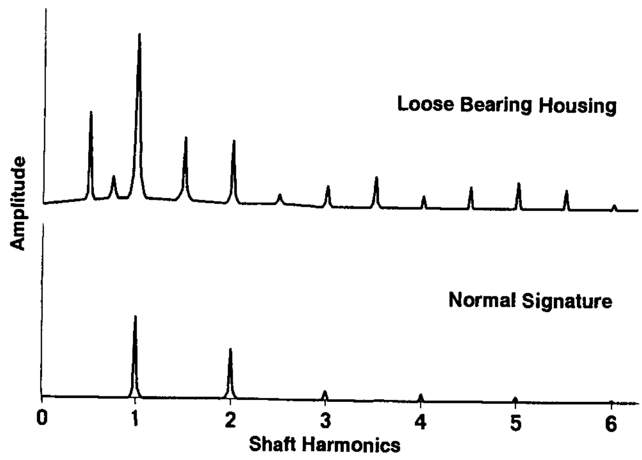


Fig. 4 Vibration characteristics of mechanical looseness.

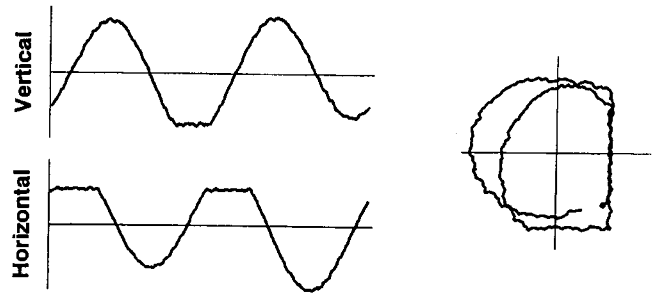


Fig. 5 Waveform and orbit of a rub.

2.5 Whirling in Rotating Machinery

One frequent failure mode in rotating machinery is caused by whirling or whipping of the shaft due to self-excited vibrations. The energy for this excitation is provided by a uniform source of power associated with the machinery system itself. This source of power can give rise to oscillating forces from various mechanisms inherent in the system. The rotor system will vibrate at its own natural or critical frequency, essentially independent of any external stimulus. The direction of whirl may be in the same direction as the shaft rotation (forward whirl) or opposite to it (backward or reverse whirl). The direction of the whirl is dependent on the direction of the destabilizing force.[3]

In rotating machines, whirling is caused by the generation of a tangential force, normal to an arbitrary radial deflection of a rotating shaft, whose magnitude is proportional to the deflection. At the onset of whirl, such a force will overcome any stabilizing external damping forces that are generally present. Whirling motions are produced with an ever-increasing amplitude, limited only by nonlinearities that ultimately limit the maximum shaft deflection. All the self-excited systems involve friction or fluid energy mechanisms in the generation of destabilizing forces. A simple mathematical analysis shows that the trajectory of the center of gravity of a whirling rotor is an exponential spiral (Fig. 6). Some important examples of whirling and whirling instabilities are hysteretic whirl, fluids trapped in hollow rotors, dry friction whip, and fluid bearing whip.

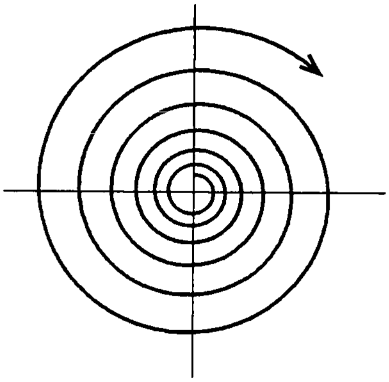


Fig. 6 Trajectory of the center of gravity of a rotor's mass due to whirling.

2.6 Hysteretic Whirl

Hysteretic whirl is caused by internal friction between the fibers within the rotating shaft or between fitted components, producing a phase shift between the neutral stress axis and the neutral strain axis. A displacement in angular orientation between the two axes gives rise to a resultant force, not parallel to the deflection of the shaft. The net effect is that the resultant force has a tangential component normal to the deflection, a prerequisite for whirling to occur. The tangential component is in the direction of rotation, inducing a forward whirl motion in the direction of rotation. This increases the centrifugal forces on the deflected rotor, thereby increasing the deflection. It has generally been recognized that hysteretic whirl can only occur at rotational speeds above the first shaft critical speed. Once whirl has started, the critical whirl speed that will be induced will have a frequency approximately half that of the onset speed.

2.7 Whirl due to Fluid Trapped in Rotor

This occurs in hollow rotors, mostly in high-speed rotating machinery, where liquids (such as oil in bearing sumps or steam condensates) may inadvertently be trapped in the internal cavity of the rotors. The spinning surface of the cavity drags fluid in the direction of rotation. The deflection of the shaft and the centrifugal forces on the trapped fluid produce a force having a tangential component, inducing a forward whirl. It has been shown that the onset speed for instability is always above the critical speed of the rotor and below twice critical speed.

2.8 Dry Friction Whip

Dry friction whip is experienced when the surface of the rotating shaft comes in contact with an unlubricated stationary surface. This can be caused by inadequate journal bearing lubrication, contact between rotors and labyrinth seals, or turbomachinery blade rubs. The contact between the surface of the rotating shaft and the stationary surface will induce a tangential force on the rotor, which then produces a whirling motion. This generates a larger centrifugal force on the rotor. The whirl produced is counter to the direction of shaft rotational (backward whirl). No first-order interdependence of whirl speed with rotational speed is established. It has been suggested that the whirl frequency is half the rotational speed of the rotor.

2.9 Fluid Bearing Whip

This effect appears in journal bearings, such as those found in high-speed turbomachinery. Radial deflection of a shaft rotating in a fluid-filled clearance entrains viscous fluid, which will circulate with an average velocity of about half the shaft speed. Pressure on the upstream side of the reduced clearance zone will be higher than that on the downstream side. The resultant bearing force will exert a tangential force component in the direction of rotation, inducing forward whirl. Any induced whirl produces increased centrifugal forces, which further reduce clearances and generates ever-increasing destabilizing tangential forces. This instability normally occurs at shaft speeds greater than twice the critical speed, with a whirling frequency at approximately half the shaft speed.

2.10 Rolling Element Bearings

Faults in a rolling element bearing start as a single defect caused by metal fatigue in either the raceways or the rolling elements. A normal bearing in good condition produces vibration signals that are essentially random, with only a few minor impulsive events present. The vibration signature of a damaged bearing is dominated by impulsive events related to ball or roller passing frequencies.[4,5] Figure 7 shows the general characteristics of a vibration signal at various stages of damage, in both the time and frequency domains. As damage spreads around the bearing, significant increases in the amplitude of both the broadband noise and the defect frequencies occur. Formulas for rolling element, inner and outer race ball passing, and train frequencies can be found in Ref. 6.

In a simple machine where the signal from the other components does not contaminate or mask the vibration signal, bearing condition can be monitored by measuring the crest factor. This is the ratio of the peak to rms values of the wide-band vibration signal and can be obtained from an accelerometer mounted directly on the bearing. Defects in the bearing components produce sharp peaks

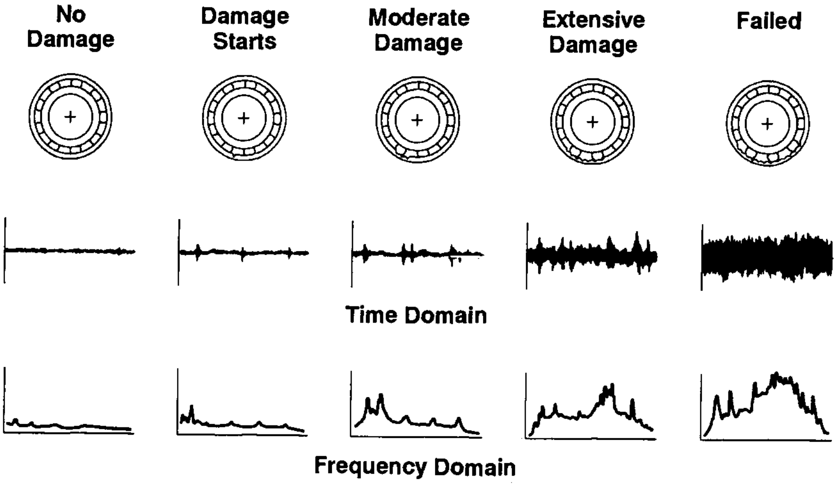


Fig. 7 Vibration characteristics of bearing at various stages of damage.

with higher crest factors. A crest factor of 4–6 is indicative of normal operation; a range of 6–12 marks the presence of damage.

The impulsive nature of the vibration signature of a defective bearing can be quantified using the normalized fourth moment of the time-domain signal. This statistical value is commonly called the kurtosis. A formula for calculating the kurtosis is given in the next section. A bearing in good operating condition will produce a kurtosis value of around 3 (indicative of random noise). Minor damage, such as a single fatigue spall, increases the impulsiveness of the signal and produces a kurtosis of about 6. As the damage spreads, the signal becomes less coherent and the kurtosis value falls. Finally, the damage becomes extensive and spreads around the periphery of the loaded region. At this stage, the vibration signature again becomes essentially random in nature because of the large number of impulsive events present, and the kurtosis values fall back to around 3. A kurtosis value of 3, for both a good and a damaged bearing, appears rather confusing at first. However, a significant increase in energy levels, as shown in Figure 7, is associated with advanced stages of damage.

Most bearing monitoring techniques rely heavily on the ability to place the transducer on the bearing housing, close to the loaded zone of the stationery outer race. The measured signal should be free from other sources of vibrations, otherwise the low-energy level of the bearing signal will be masked by background vibration signals. Using advanced signal processing techniques, the bearing signal can be recovered from the overall signal of the machinery system. One of the techniques used is called the high-frequency resonance technique and is described in the next section.

2.11 Gears

Vibration analysis is a powerful technique for the early detection of failure in gearboxes and transmissions.[7] A nominally perfect gear produces almost sinusoidal signals in the time domain, with prominent peaks at the tooth meshing frequency and lower-order harmonics in the frequency domain. Faults, such as a localized tooth defect or misalignment, alter the signal in the time domain and change the pattern of the frequency spectrum, yielding diagnostic information. Common faults that occur in gears can be identified from the time- and frequency-domain information. In a complex gearbox, the vibration signature obtained from an accelerometer placed on the outside surface contains contributions from all the components of the system. The spectrum of the overall signature cannot always give diagnostic information, as the signature of the gear of interest is often masked by the signatures of other components. Using synchronous time-domain averaging, explained in the next section, the time waveform of individual gears can be recovered from the overall signature.

Time-domain averaged waveforms and frequency spectra for some common gears faults are shown in Fig. 8.[8] A nominally perfect gear exhibits a sinusoidal waveform in the time domain, with a dominant peak in the spectrum at the meshing frequency. A misaligned gear produces a once-per-rev amplitude modulation in the time domain. A broken or fractured gear tooth fails to transmit its share of the load; the gears accelerate to engage the next pair of teeth and produce an impact upon engagement. This impulsive behavior produces sidebands about the mesh frequency and its harmonics in the frequency domain. Spalled teeth and soft

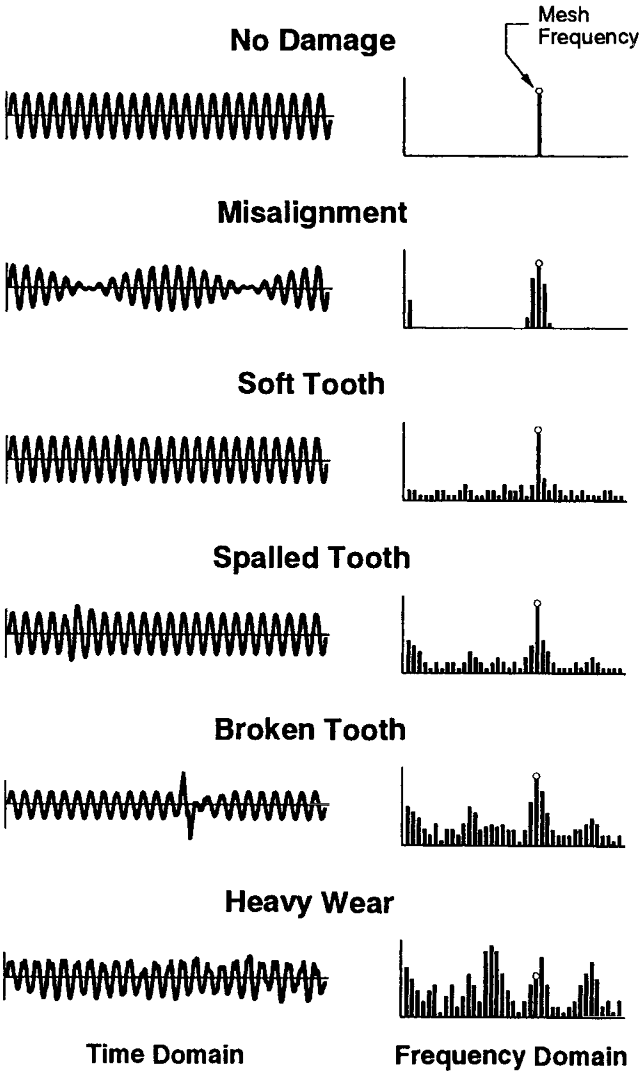


Fig. 8 Waveforms and spectra produced by common gear faults.

teeth behaved in a similar manner. A heavily worn gear, with wear all around or on a large number of teeth, produces a signature that is quite random in character.

Faults in gears can also be detected by the application of amplitude and phase modulation analysis techniques on the time-domain-averaged vibration signal. Signal modulations produced by a defective gear can be analyzed in the time domain itself, by narrow-band filtering about the mesh frequency and harmonics.[9–11] An example of this is shown in Fig. 9.

3 SIGNAL ANALYSIS AND PROCESSING TECHNIQUES

Condition monitoring, by definition, is the act of extracting information from a specific system, by means of appropriate measurements. Vibration monitoring consists of acquiring vibration data for use as information carriers on the condition of the machinery system/component. The raw vibration data must be processed and analyzed to determine whether or not a fault exists.[12–14] It can also be important to determine the nature of the fault; for example, to know whether the fault is local tooth damage or misalignment in gear condition monitoring. The processing techniques and analysis methods frequently depend on the machinery component and the nature of the fault to be detected.

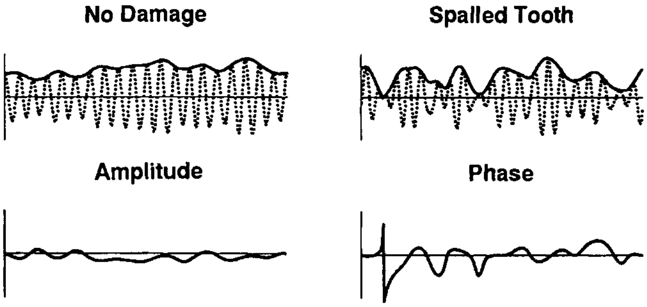


Fig. 9 Amplitude and phase time-domain signatures of undamaged and damaged gears.

3.1 Data Processing and Analysis

The basic operation of a data acquisition and processing system is described in this section. The analog signal from the sensor is routed through an antialiasing filter (low-pass filter), before performing analog-to-digital conversion. The main objective in signal processing is to enhance the desired signal and remove any undesirable features. This is done to improve the signal-to-noise ratio and facilitate the recovery of the signal generated by the machinery component of interest.

The next phase of signal processing and analysis is concerned with extracting features associated with a particular type of fault or component damage. In simple machines, these could be based on signal energy level measurements, such as rms, peak, crest factor, or spectral composition involving frequency-domain analysis. Complex machines with several components require more specific diagnostic techniques and tools to extract the features associated with particular faults. Analysis methods[15,16] use both time- and frequency-domain data, applying several advanced signal processing techniques. Pattern recognition techniques and statistical methods are also extremely useful in feature extraction. Detailed discussions on instrumentation and signal processing procedures can be found in Chapter 72 and Chapter 101.

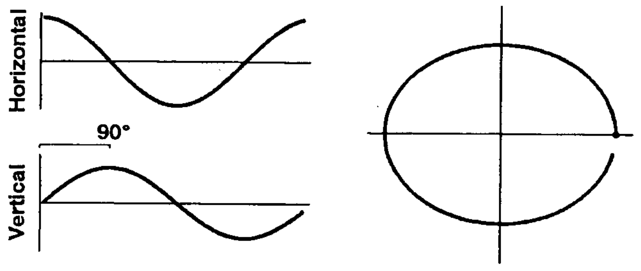
Frequency-domain analysis methods dominate most mechanical signal analysis applications. The prevalence of spectral analysis has increased even more with the emergence of digital analyzers. The potential value of time-domain methods might have been overlooked. The recognition and analysis of patterns in the time domain is useful in identifying faults related to the kinematic motion of the machinery. In the following paragraphs we will discuss some of the time- and frequency-domain methods. Commercial software packages are available for most of the analysis techniques and some have been implemented in instrumentation hardware.

3.2 Orbit Displays

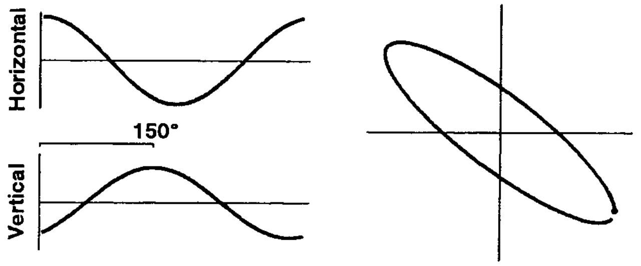
Many common problems in rotating machinery, such as imbalance and whirling, can be identified from the shape of a shaft orbit.[2] Orbit diagrams are generated using two noncontact vibration displacement transducers spaced 90° apart in the radial direction of the shaft. Various shaft orbit patterns are shown in Fig. 10. A circular orbit is obtained from two sinusoidal waveforms equal in amplitude. Two sinusoidal waveforms of different amplitudes produce an elliptical orbit, possibly indicating misaligned coupling restraints. If the two waveforms contain more than one discrete frequency, the resulting orbit pattern will increase in complexity.

3.3 Time-Domain Averaging

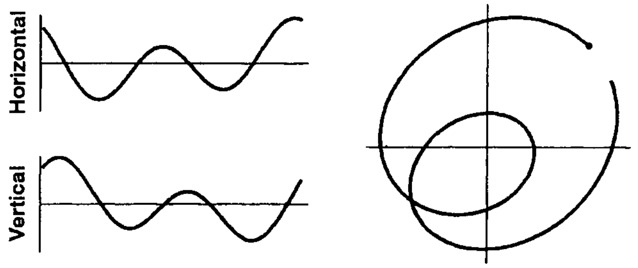
One of the advantages of time-domain averaging lies in the possibility of relating signal changes to specific kinematic events. Time-domain analysis can readily be performed on rotating machinery, as it is relatively easy to access reference time signals from sensors monitoring the main drive shaft. Synchronous time averaging with respect to the rotation of a component eliminates any unrelated external signals. Figure 11 shows how the synchronous time average is derived for one of the gears in a gearbox. The time record length of the averaged signal corresponds to one complete mesh cycle of the gear (the time taken for all of the teeth on a gear to pass through mesh).



(a) Orbit produced by two sinusoids with a phase difference of 90°.



(b) Orbit produced by two sinusoids with a phase difference of 150°.



(c) Orbit produced by two complex periodic waveforms.

Fig. 10 Orbit displays.

3.4 Amplitude-Domain Analysis

In wear-induced failures, localized defects appear first. As these grow and spread, the damage becomes more distributed in nature. The vibration signature often is impulsive in nature at the onset of a localized defect, tending to change with time to a more continuous function. The sharp peaks at the onset of a defect affect the tails of the probability density function and the moments of the distribution. The crest factor and kurtosis function are used to describe the shape characteristics of the signal. Crest factor and kurtosis are defined by

$$\text{Crest factor} = \frac{\text{peak}}{\text{rms}}, \tag{1}$$

$$\text{Kurtosis} = \frac{1}{N} \frac{\sum_{n=1}^{N} (x_n - \bar{x})^4}{\left(\sum_{n=1}^{N} (x_n - \bar{x})^2\right)^2}. \tag{2}$$

The kurtosis is sensitive to the impulsiveness or "spikiness" of the data. The kurtosis of random or Gaussian noise is 3.0. The vibration signal of a rolling element bearing in good condition is expected to have this value. The kurtosis level rises sharply to around 6 at the onset

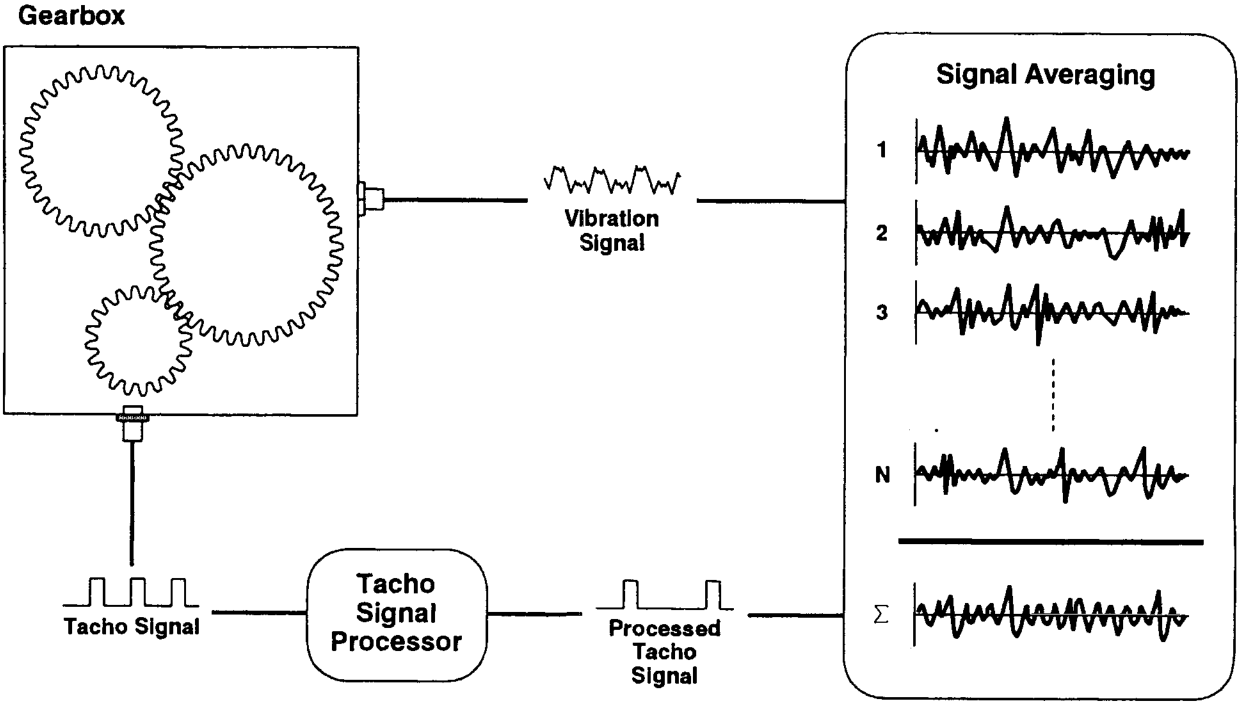


Fig. 11 Time-domain averaging for gear signal processing.

of discrete damage (i.e., a spall on the rolling elements or raceways). The kurtosis value falls as the damage spreads, indicating that the damage characteristics have become more random in nature.

3.5 Envelope Spectrum Analysis Technique

Most bearing monitoring techniques rely on the ability to place the vibration transducer on the bearing housing, close to the outer raceway load zone, to reduce contamination of the measured signal by extraneous sources of vibrations. However, it is not always possible to mount the transducer close enough to the bearing. The envelope spectrum analysis technique can be used to analyze signals from a transducer mounted at a more remote location, containing extraneous vibration signals from other machinery components. This technique makes use of high-frequency resonances that act as mechanical amplifiers and as carriers of the lower frequency fault signals. The initial process of locating a suitable resonant carrier frequency is the most difficult part of the envelope spectrum technique. Envelope analysis is not restricted to remote monitoring applications, as many bearing faults can be identified by modulations in the vibration signature.

The input signal from an accelerometer is routed through a bandpass filter centered on the resonant frequency. This produces a time-varying signal modulated by the transient generated by the defect. The filtered signal is rectified and then demodulated using an envelope detector. Alternatively, the amplitude demodulation techniques described in Section 3.6 can be employed to extract the envelope. The resulting signature is dominated by rolling element defect frequencies. The power spectrum of the envelope is then calculated. Any changes in defect-related frequency levels or patterns are assessed for fault evaluation.

3.6 Amplitude and Phase Modulation Techniques for Gear Analysis

A fault in any rotating machine is likely to introduce sidebands, due to amplitude and frequency modulations, around prominent frequency bands associated with the machine. In a simple pair of gears, it is relatively easy to identify sidebands around tooth meshing frequencies. In complex gear systems comprising large numbers of gear sets (e.g., helicopter transmission systems), it is more difficult to identify any specific gear because of the large number of meshing frequencies and sidebands present.

The vibrations produced by gears can be characterized as the normal gear meshing vibrations, plus additional effects producing sidebands around the mesh frequency harmonics. Defects can be identified by applying amplitude and phase modulation analysis techniques. These were originally proposed by McFadden and Smith[9] and further developed by Nicks and Krishnappa.[10, 11] Under ideal conditions, the time-domain average of the vibration of a pair of gears is a periodic function that consists solely of tones at the fundamental and harmonics of the mesh frequency. A discrete gear fault, such as spall or a fatigue crack, modifies this equation by introducing periodic phase or amplitude modulations in the vibration signal. The equation for the modulated vibration signal can be written as

$$y(t) = \sum_{m=0}^{M} X_m[1 + a_m(t)] \cos[2\pi m T f t + \theta_m + b_m(t)], \quad (3)$$

where X_m is the vibration amplitude at the mesh frequency harmonic m $(0, 1, 2, \ldots, M)$, $a_m(t)$ is the amplitude modulation function, T is the number of teeth on the gear, f is the rotational frequency, θ_m is the phase lag, and $b_m(t)$ is the phase modulation function.

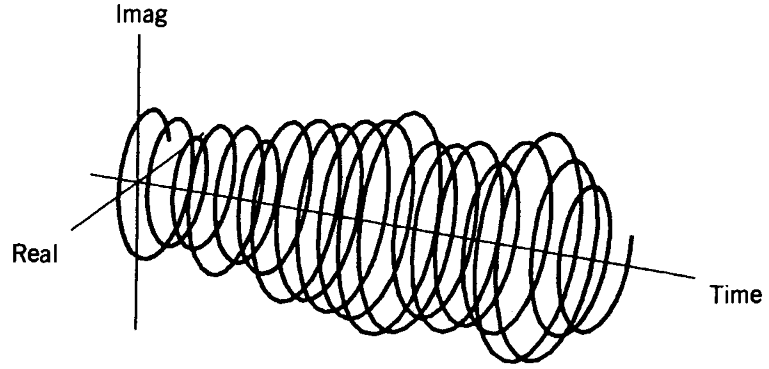
In the frequency domain, the Fourier transform of the above function comprises the fundamental and harmonics of the meshing frequency, surrounded by modulation sidebands. The function $y_m(t)$ can be approximated by bandpass filtering the time-domain averaged vibration signal. The amplitude envelope and phase modulation trace are extracted from the complex analytical function $z_m(t)$, given by

$$z_m(t) = y_m(t) + jH[y_m(t)], \quad (4)$$

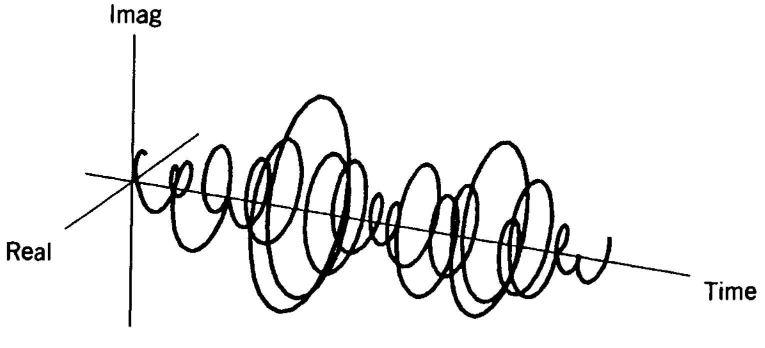
where H is the Hilbert transform of $y(t)$. Examples of this are shown in Fig. 12.

3.7 Cepstrum Analysis

The application of cepstrum analysis to machinery diagnostics is mainly based on its ability to detect periodicity in the spectrum. In simple terms, the cepstrum can be considered to be the spectrum of a spectrum. It was originally defined to be the power spectrum of the logarithm of the power spectrum. The terminology used in cepstrum analysis is derived from that of frequency analysis: spectrum ⇔ cepstrum, frequency ⇔ quefrency, and harmonics ⇔ rahmonics.



(*a*) Gear in good condition.



(*b*) Surface-damaged gear.

Fig. 12 Complex analytical vibration signatures of gears.

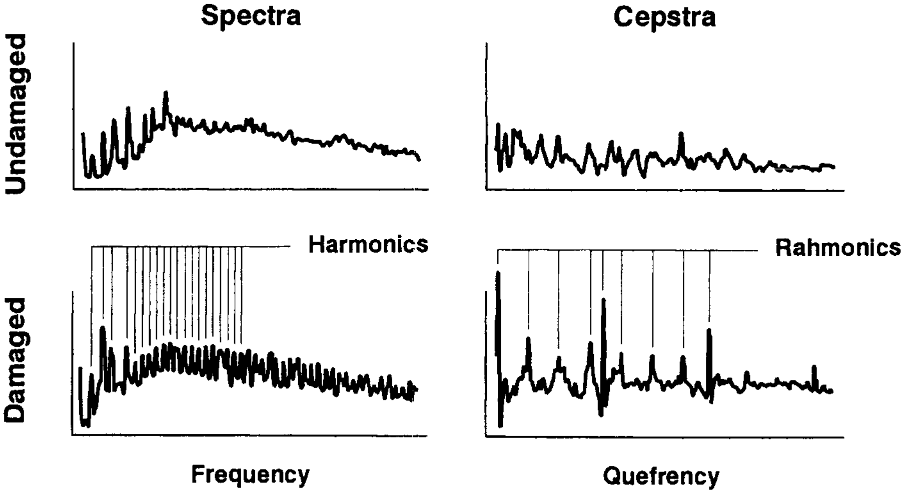


Fig. 13 Spectra and cepstra of undamaged and damaged gearboxes.

Cepstrum analysis can be used as a tool for the detection of families of harmonics with uniform spacing. The logarithmic amplitude scale emphasizes the harmonic structure of the spectrum and reduces the influence of the transmission path by which the signal travels from the source to the measurement point. Faults of various kinds in a gearbox are indicated by groups of harmonics and their sidebands. Although the frequency patterns can be seen by the eye, they may be quite complex and difficult to interpret. Cepstrum analysis is used to reduce these groups to a few rahmonics that can be much easier to monitor. A more detailed description of cepstrum analysis can be found in Ref. 17.

The cepstrum is now defined as the inverse transform of the logarithm of the power spectrum, or

$$C(\tau) = \mathcal{F}^{-1}(\log |X(f)|^2), \qquad (5)$$

where $X(f)$ is the Fourier transform of a time series $x(t)$. One of the reasons for this new definition is to highlight the connection between the cepstrum and the autocorrelation function, defined as the inverse transform of the power spectrum:

$$R_x(\tau) = \mathcal{F}^{-1}(|X(f)|^2). \qquad (6)$$

The independent variable τ of the cepstrum has the dimensions of time but is known as quefrency. A high quefrency indicates rapid fluctuations in the spectrum (small frequency spacings), and a low quefrency indicates slow changes with frequency (large spacings). The quefrency of the peak represents the periodic time of modulations in the time domain (i.e., sideband spacing in the frequency domain). The quefrency only identifies frequency-domain spacing—it tells nothing about absolute frequencies.

Spectra and cepstra for two gearboxes are shown in Fig. 13. The spectrum of the gearbox in good condition does not show any significant periodicity, and this is reflected in the cepstrum where few discrete rahmonics are present. In the damaged gearbox, however, a large number of discrete peaks are present in the spectrum. The cepstrum clearly shows that these are periodic and identifies the fundamental period.

REFERENCES

1. E. L. Thearle, "Dynamic Balancing of Rotating Machinery in the Field," *Trans. ASME*, Vol. 56, No. 10, 1934, pp. 745–753.
2. J. S. Mitchell, *An Introduction to Machinery Analysis and Monitoring*, Penwell, 1981, pp. 107–204.
3. C. M. Harris and C. E. Crede, *Shock and Vibration Handbook*, 2nd ed., McGraw-Hill, New York, 1976, pp. 5.1–5.24.
4. R. H. Bannister, "A Review of Rolling Element Bearing Monitoring Techniques, I. Mech. E. Condition Monitoring Machinery Plant," *Symposium of Condition Monitoring Machinery Plant*, 1985, pp. 11–24.
5. N. S. Swanson and S. C. Favalord, *Applications of Vibration Analysis to the Condition Monitoring of Rolling Element Bearings*, Aeronautical Research Laboratory, Aero Propulsion Report 163, 1984.
6. P. Eschman et al., *Ball and Roller Bearings*, Wiley, 2nd ed., 1985, pp. 78–138.
7. J. D. Smith, *Gears and Their Vibration*, Marcel Dekker, New York, 1983.

8. R. M. Stewart, "Some Useful Data Analysis Techniques for Gearbox Diagnostics," *Proceedings of the Symposium "Applications of Signal Processing,"* Institute of Sound and Vibration Research, Southampton, UK.
9. P. D. McFadden and J. D. Smith, "A Signal Processing Technique for Detecting Local Defects in a Gear from the Signal Average of the Vibration," *Proceedings of the I. Mech. E.*, Part C, Vol. 199, 1985, pp. 287–292.
10. J. E. Nicks and G. Krishnappa, "A Preliminary Study into the Use of Phase Demodulation Techniques for the Analysis of Gear Vibration Data," National Research Council, Ottawa, TR-ENG-005, 1989, NRC Report No. 32083.
11. J. E. Nicks G. and Krishnappa, "Amplitude and Phase Modulation Analysis Methods for Early Detection of Gear Faults," *Proceedings of the International Congress on Recent Developments in Air- and Structure-Borne Sound and Vibration*, Auburn, Alabama, 1990, pp. 263–270.
12. R. H. Lyon, *Machinery Noise and Diagnostics*, Butterworth, Stoneham, MA, 1987, pp. 17–47.
13. S. Braun, *Mechanical Signature Analysis, Theory and Applications*, Academic, London, 1986.
14. C. Cempel and H. D. Haddad, *Vibroacoustic Condition Monitoring*, Ellis Horwood, 1991.
15. J. S. Bendat and A. G. Piersol, *Random Data Analysis and Measurement Procedures*, Wiley-Interscience, New York, 1971.
16. E. O. Brigham, *The Fast Fourier Transform and its Applications*, Prentice-Hall, Englewood Cliffs, NJ, 1988.
17. R. B. Randall and J. Hee, Cepstrum Analysis, *Brüel and KjærTech. Rev.*, No. 3, 1981, pp. 1–32.

74

STRUCTURE-BORNE ENERGY FLOW

G. Pavic

1 INTRODUCTION

1.1 Structural Intensity and Energy Flow

In an elastic body, structural vibration induces the flow of mechanical energy through the body. The energy flow (EF) results from the interaction between vibration-generated stresses and movements throughout the body. The EF at any single point represents the instantaneous rate of energy transfer per unit area in a given direction and is called the structural intensity (SI). Defined in such a way, the intensity becomes a vectorial quantity associated with the observed point. Its physical unit is watts per metre squared. When integrating intensity across an area within the body, the total mechanical energy flow through this area is obtained. The unit of energy flow is the watt.

The intensity vector changes both its magnitude and direction in time. Variations of intensity at different points within the structure need not coincide mutually. In order to compare energy flows at various locations in a meaningful way, their time-averaged (net) value must be analyzed. Thus, the intensity concept is best associated with vibration fields that exhibit temporal stationarity.

Once the vibrational energy flow through a structure is known, it becomes possible to identify the positions of vibratory sources, propagation paths, and absorption areas. The nature of this information, which is not present in data on vibration levels, implies that energy flow techniques should be used primarily in the fields of source characterization, system identification, and vibration diagnostics.

The SI or EF of structures can be examined either by measurement or by computation, depending on the nature of the technical problem being considered. Measurements are used for diagnostic or source identification purposes, while computation is suitable for noise and vibration control during the design stage.

1.2 Energy Flow Concepts

The characteristics of energy transfer by structural vibration are such that three complementary concepts are viable in representing vibration-induced EF:

(a) The direct intensity concept utilizes SI (energy flow per unit area) as the final result in an analysis. This concept is completely analogous to sound intensity, as it represents the same physical quantity. The intensity concept is a tool for three-dimensional visualization of energy flow within an elastic body exhibiting structural vibration. It is more suitable for computational than measurement work, as any measurement of practical use made on an object of complex shape has to be restricted to the exterior surfaces only. The latter condition leads to a specific quantity, the surface structural intensity (SSI). In the majority of cases, surface intensity values can be used to qualify rather than to quantify the energy flow within the object under examination.

(b) Structures of uniform thickness (plates, shells) can be more appropriately analyzed if the intensity is integrated across the thickness. In this way a unit energy flow (UEF) is obtained, that is, the flow of energy per unit length normal to any line on the surface. If the energy exchange between the outer surfaces and the surrounding medium is small in comparison with the flow tangential to the surface, as it is for objects in contact with air, then the UEF vector lies in the tangential plane of the middle surface. For thin-walled structures, where

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker
Copyright © 1997 John Wiley & Sons, Inc.

the wavelengths are much larger than the thickness, the outer surface motions are related in a simple way to those in the interior (Kirchhoff's assumptions). This fact makes it possible to express UEF in terms of outer surface quantities, which in turn makes it possible to measure it directly. The UEF concept is best utilized for the needs of source localization and fault diagnostics.

(c) Structural parts that form waveguides or simple structural joints can most appropriately be analyzed by means of the concept of total energy flow (TEF). Some such parts, such as beams and elastic mountings, are of a simple enough shape to allow expression of the EF by simple formulas in terms of vibrational displacements. The TEF in waveguides of a more complex mechanical character, such as pipes filled with fluid, can then be evaluated by more complex procedures such as the wave decomposition technique.

2 PHYSICAL BACKGROUND OF STRUCTURAL ENERGY FLOW

Expressed in Cartesian coordinates, the three components of an SI vector read

$$\begin{aligned} I_x &= -\sigma_{xx}\dot{u}_x - \sigma_{xy}\dot{u}_y - \sigma_{xz}\dot{u}_z, \\ I_y &= -\sigma_{yx}\dot{u}_x - \sigma_{yy}\dot{u}_y - \sigma_{yz}\dot{u}_z, \\ I_z &= -\sigma_{zx}\dot{u}_x - \sigma_{zy}\dot{u}_y - \sigma_{zz}\dot{u}_z. \end{aligned} \tag{1}$$

Here σ are dynamic stresses and u the particle displacements at the observation point, their subscripts indicating component orientation. The dots above the symbols for displacements indicate time derivative; that is, intensity components are products of stresses and particle velocities.

Since the shear stress components obey symmetry, for example, $\sigma_{xy} = \sigma_{yx}$, the complete SI vector is built up of nine different quantities, six stresses and three velocities. Each of the SI components is a time-varying quantity. Therefore SI is a time-varying vector. Its magnitude and direction both change in time. It should be pointed out that the SI vector does not necessarily have to be normal to the wavefront surface.

The SI formulas (1) are similar to those for sound intensity components but exhibit two important differences. First, the SI formulas are more complex, since each component consists of three terms instead of a single one, as in the case of sound intensity. In gases and liquids the shear stresses due to sound motions are very small and can be disregarded when compared with normal stresses. Another difference is in the sign of each product, which merely results from the sign convention. Therefore, sound intensity is simply a special case of Eqs. (1) when shear effects vanish.

At a free surface, the intensity formulas (1) become simpler. If the coordinate z is taken as the normal to the surface, then all z-dependent terms vanish, reducing thus the intensity vector to I_x and I_y components each containing two terms only.

To determine the SI or EF values in a particular case, time-varying components of both the stresses and the velocities have to be known at the observed point. Stresses are not measurable in a direct way; in order to determine them, relationships are used that link them to (measurable) strains. For homogeneous and isotropic materials behaving linearly, such as metals, most plastics, rubber, and so on, the stress–strain relationships are fairly simple. These relationships contain material constants: Young's and shear moduli.[1]

Equations (1) refer to instantaneous values of intensity, the knowledge of which is not useful for most analytical purposes. Similarly, instantaneous values of sound pressure or vibration velocity are of no particular analytical use. It is therefore desirable to operate with averaged values of intensity and energy flow.

Time-averaged SI represents net energy flow per unit surface area at the observed point. It is termed "active" intensity, because it refers to the portion of total energy flow taken away from a given region without being returned. The remaining portion of the energy flow, which fluctuates to and fro but has zero mean value, is accordingly termed "reactive" energy.

3 FORMULAS FOR ENERGY FLOW

The basic SI formulas (1) are not sufficient for measurement of either unit or total energy flow. Using these formulas, the expressions for UEF or TEF can be evaluated for some simpler type of structures: rods, beams, plates, and shells. As a rule, a different governing formula will apply to each type of structure.

In this section formulas for EF are listed. The formulas are given in terms of strains and velocities, that is, the measurable quantities. Both the strains and the velocities can be detected by various techniques, but only at the outer surface of the object under investigation. In order to determine these quantities within the body, some knowledge is required of the internal velocity and strain distribution. For certain structures (rods, beams, plates, and shells) very simple relationships are sufficient to express internal quantities in terms of the external ones, as long as the wavelengths of the structural motion are much larger than the lateral dimensions of the object.

3.1 Longitudinally Vibrating Rod

In a longitudinally vibrating rod, only the axial component of the stress differs from zero; thus only the axial normal strain exists. The SI distribution in the cross section is uniform (Fig. 1*a*):

$$I_x = -E\varepsilon_{xx}\dot{u}_x. \tag{2}$$

The total EF is obtained by multiplying the intensity with the area of the cross section S:

$$P_x = I_x S. \tag{2a}$$

3.2 Flexurally Vibrating Beam

In this case the intensity varies over the cross section due to variations of both the stresses and the velocities caused by flexion (Fig. 1*b*). Two components of stress exist: the axial normal stress σ_{xx} and the lateral shear stress σ_{xz}. The axial stress varies linearly with distance from the neutral plane, while the variation of the shear stress depends on the shape of the cross section. For a rectangular cross section this variation is parabolic with the maximum value in the middle plane. The axial component of particle velocity, $\dot{u}_x$, exhibits the same form of variation as the axial stress, while the lateral component, $\dot{u}_z$, remains constant. Integration of axial intensity across the lateral section leads to expressions for the total EF. Since the axial and lateral displacements are coupled through a simple expression $u_x = -z\,\partial u_z/\partial x$, various formulations of the EF are possible:

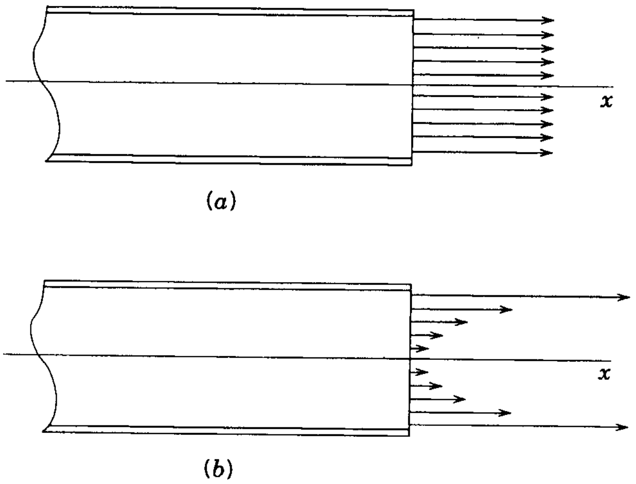


Fig. 1 Intensity distribution in a rod (beam): (*a*) longitudinal; (*b*) flexural vibration.

$$\begin{aligned} P_x &= JE\left(\frac{\partial^3 u_z}{\partial x^3}\,\dot{u}_z - \frac{\partial^2 u_z}{\partial x^2}\,\frac{\partial \dot{u}_z}{\partial x}\right) \\ &= \frac{JE}{\delta^2}\left(\delta\,\frac{\partial \varepsilon_{xx}}{\partial x}\,\dot{u}_z + \varepsilon_{xx}\dot{u}_x\right), \end{aligned} \tag{3}$$

where J is the area moment of inertia about the bending axis and δ the distance from the neutral plane to the outer surface, which the quantities in Eq. (3) refer to. The first formulation[2] contains only lateral motions, displacements, and velocities, while the second one contains strains and velocities.[3]

3.3 Torsionally Vibrating Bar

Torsion in a bar produces shear stresses that rise linearly from the center of the bar. If the bar is of circular or annular cross section, these shear stress components are the only nonvanishing ones. In such a case the EF through the bar, obtained by integrating the SI through the cross section, reads (Fig. 2)

$$P_x = -T\,\frac{\partial\varphi}{\partial x}\,\dot{\varphi} = -\frac{T}{R_0}\,\varepsilon_{x\varphi}\dot{\varphi}. \tag{4}$$

Here φ is the angular displacement of the bar, $\varepsilon_{x\varphi}$ is the shear strain at the outer radius R_0 and T is the torsional stiffness. For a bar of an annular cross section, the torsional stiffness is equal to $T = \pi/2G(R_0^4 - R_1^4)$.

The torsional stiffness of a solid shaft is obtained by setting the inner radius R_1 to zero. If the bar is of rotationally nonsymmetrical cross section, Eq. (6) applies in the same way but in addition axial motions and stresses

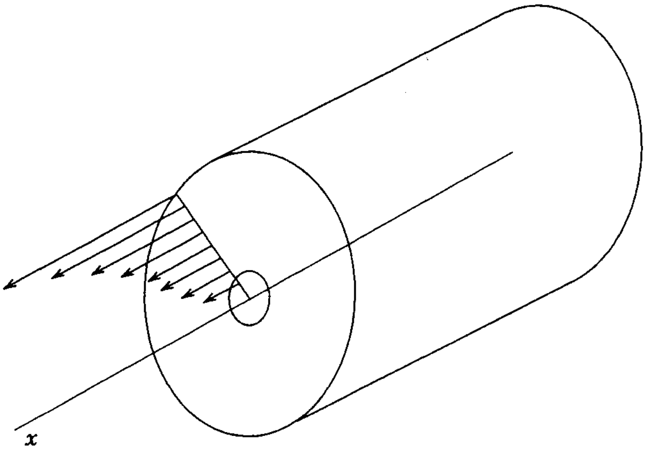


Fig. 2 Intensity distribution in a torsionally vibrating bar.

take place that can generate the flow of energy. The mechanism of such an additional EF very much depends on the shape of the cross section and should be analyzed separately for each individual case.

3.4 Longitudinally Vibrating Thin Plate

For a thin plate exhibiting in-plane motion only, the intensity is constant throughout the thickness. The SI component in the x-direction reads[3] (Fig. 3)

$$\begin{aligned} I_x &= -E(1-\nu^2)[\varepsilon_{xx}\dot{u}_x + \tfrac{1}{2}(1-\nu)\varepsilon_{xy}\dot{u}_y], \\ P'_x &= I_x h, \end{aligned} \tag{5}$$

while for the y-direction an analogous expression applies, obtained by interchanging the x and y subscripts. The UEF P' is simply obtained by multiplying the intensity with the plate thickness h. The expression is similar to that for the intensity in the rod [Eq. (2a)], the difference being an additional term due to shear stresses carrying out work on motions perpendicular to the observed direction.

3.5 Flexurally Vibrating Thin Plate

As for the flexurally vibrating beam, the normal component of vibration in a flexurally vibrating plate is of a much higher level than the in-plane component. It is therefore appropriate to formulate the intensity expressions in terms of the normal component u_z. With respect to intensity distribution throughout the cross section, a situation applies analogous to that for the beam, with the exception of an additional mechanism of energy flow generation, due to twisting deformations of the plate. The UEF in the x-direction reads[2,4]

$$\begin{aligned} P'_x = \frac{Eh^3}{12(1-\nu^2)} \Bigg[& \frac{\partial(\Delta u_z)}{\partial x}\,\dot{u}_z - \left(\frac{\partial^2 u_z}{\partial x^2} + \nu\frac{\partial^2 u_z}{\partial y^2}\right) \\ & \cdot \frac{\partial \dot{u}_z}{\partial x} - (1-\nu)\,\frac{\partial^2 u_z}{\partial x \partial y}\,\frac{\partial \dot{u}_z}{\partial y}\Bigg], \end{aligned} \tag{6}$$

where Δ is the Laplacian. The UEF in the y-direction is obtained by x–y subscript interchange. Both components of the in-plane motion are related to normal motion, which makes possible an alternative expression for the UEF:

$$\begin{aligned} P'_x = -\frac{Eh}{3(1-\nu^2)} \cdot \Bigg[& \frac{h}{2}\,\frac{\partial(\varepsilon_{xx}+\varepsilon_{yy})}{\partial x}\,\dot{u}_z \\ & + (\varepsilon_{xx} + \nu\varepsilon_{yy})\dot{u}_x + \frac{1-\nu}{2}\,\varepsilon_{xy}\,\dot{u}_y\Bigg]. \end{aligned} \tag{6a}$$

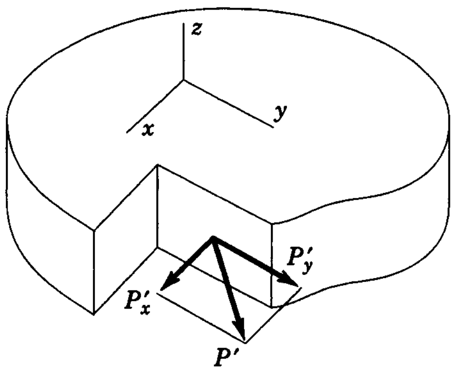


Fig. 3 Energy flow in a thin plate.

3.6 Composite Vibrations of Beams and Plates

In a beam (rod, bar) all three types of vibration discussed in Sections 3.1–3.3 can exist simultaneously. The same applies to longitudinal and flexural vibrations in a flat plate. In both cases, the different types of vibration are uncoupled, and consequently the intensities due to each of the different types of vibration should be summed algebraically. The problem here is that both longitudinal and flexural vibrations produce longitudinal motions. In order to apply intensity formulas correctly to a measurement, the origins of the motions must be identified. This can be done by connecting the measuring transducers in such a way as to suppress the effects of a particular type of motion. It should be borne in mind that not only the longitudinal vibration velocities but also the longitudinal strains are affected by flexural motion. The following formulas are used to extract the longitudinal velocity and strain components (denoted by subscript zero) from the total velocity and strain at the outer surface:

$$\begin{aligned} &\dot{u}_{x0} = \dot{u}_x + \frac{h}{2}\,\frac{\partial \dot{u}_z}{\partial x}, \qquad \varepsilon_{xx0} = \varepsilon_{xx} + \frac{h}{2}\,\frac{\partial^2 \dot{u}_z}{\partial x^2}, \\ &\varepsilon_{xy0} = \varepsilon_{xy} + h\,\frac{\partial^2 \dot{u}_z}{\partial x \partial y}. \end{aligned} \tag{7}$$

Again, components of velocity and strain in the y-direction are obtained from Eq. (7) by interchanging the x and y subscripts. It can be seen that measurements of EF due to longitudinal vibrations in the presence of flexural vibrations necessitates correction of the governing formulas (2) and (5) according to Eq. (7).

3.7 Simplified Governing Formulas for Flexurally Vibrating Beams and Plates

In regions far from vibration sources, such as discontinuities and boundaries, where the vibration field is essentially a superposition of plane waves, a simplified formula applies for net EF (TEF) due to flexural motion[5]:

$$P_x = C\omega \left\langle u_z \frac{\partial \dot{u}_z}{\partial x} \right\rangle = -\frac{C}{\omega} \left\langle \ddot{u}_z \frac{\partial \dot{u}_z}{\partial x} \right\rangle , \tag{8}$$

where

$$C = \begin{cases} h^2\rho c/\sqrt{3} & \text{(plate)}, \\ 2\sqrt{JS}\rho c & \text{(beam)}. \end{cases}$$

The constants ρ and c represent mass density and longitudinal wave velocity, respectively; ω is the radian frequency.

3.8 Thin Circular Cylindrical Shells

Due to the curvature of the shell, the in-plane and the normal motions are coupled.[6] The expressions for the UEF along the wall of the shell are the same as for the flat plate, with the addition of a UEF term P'_c to take account of the curvature:[7]

$$P'_{\text{shell}} = (P'_{\text{longitudinal}} + P'_{\text{flexure}})_{\text{plate}} + P'_c. \tag{9}$$

The two terms in parentheses are given by Eqs. (5) and (6). The curvature term is fairly complex; however, for thin shells it reduces to a simple form.[8] If the shell is positioned axially in the x-direction (Fig. 4), with the local x–y–z coordinate system attached to the observation point, the axial and tangential components of the simplified curvature term read

$$P'_{cx} \approx -\frac{\nu}{a} u_z\dot{u}_x, \qquad P'_{cy} \approx -\frac{1}{a} u_z\dot{u}_y. \tag{9a}$$

Here, as in the case of composite plate vibrations, the in-plane (longitudinal) motions due to bending must be separated from the extensional motions before expressions (5) and (9a) can be applied. Corrections for the effect of bending [Eq. (7)], which affects longitudinal motion measured at the outer surface, apply for thin shells in the same way as for thin plates.

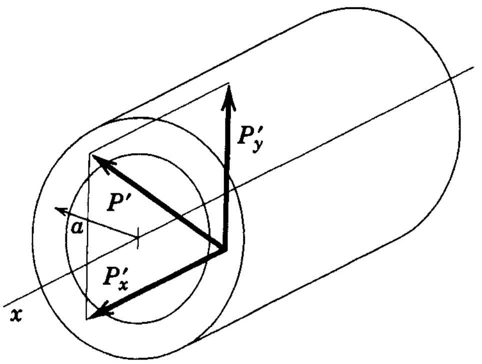


Fig. 4 Energy flow in a thin cylindrical shell.

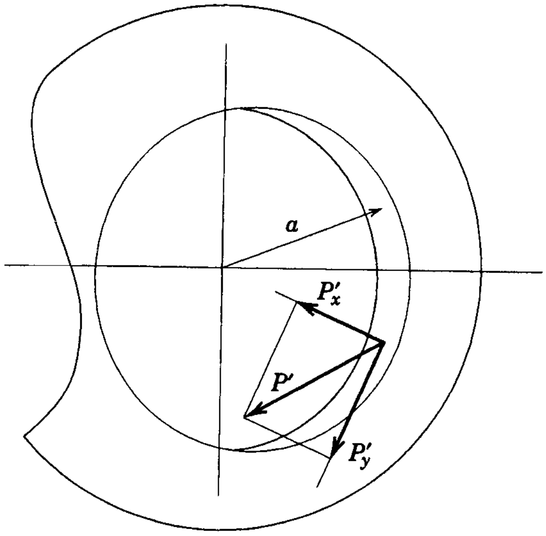


Fig. 5 Energy flow in a thin spherical shell.

3.9 Thin Spherical Shells

Figure 5 shows a section of a spherical shell with the local Cartesian coordinate system fixed to the observation point. The same UEF expression applies as for the cylindrical circular shell, Eq. (9). The only difference is in the curvature-dependent term, which is approximately equal to[9]

$$P'_x \approx -\frac{1+\nu}{a} u_z\dot{u}_x. \tag{10}$$

In this case there are no differences in the form of the x- and y-components of the curvature terms, as was the case for the cylindrical shell. This results from the properties of a spherical shell, which are centrally symmetrical about the normal to the surface.

3.10 Pipes

At not too high frequencies where higher order cylindrical waves cannot yet propagate, three types of pipe vibration occur: longitudinal, torsional, and flexural. The limiting frequency corresponding to such conditions is given by

$$f_{\text{lim}} \approx \frac{f_{\text{ring}}(6h/d)}{\sqrt{15+12\mu}}, \qquad h \ll d, \tag{11}$$

where $f_{\text{ring}} = c/\pi d$ is the pipe ring frequency, μ is the mass ratio of the contained fluid and the pipe wall, c is the velocity of longitudinal waves in the wall, and d is the mean diameter.

It turns out that the three types of motion contribute independently their respective energies to the total energy flow.[6] This enables straightforward measurements.[10,11] These three contributions should be determined separately using the formulas in Sections 3.1–3.3 and then added together.

3.11 Resilient Elements

A typical resilient element, used for vibro-isolation, consists of an elastic layer between two rigid surfaces. The forces acting at the two sides of the element stand in a firm relationship with the resulting velocities at the terminal points. This relationship is frequency dependent and is usually expressed in terms of the direct and transfer impedances of each of the endpoints.[12] Once the force–velocity relationship has been determined, either by computation or by measurement, knowledge of the end velocities of the mounted element is sufficient for evaluation of the TEF throughout the element. Due to inevitable internal losses in the element, the input EF must be larger than the output EF.

A simplified diagram of a resilent element is given in Fig. 6.

On the assumption of pure translational motion in the direction of its axis, the element is characterized by four impedances, Z_{11}, Z_{22}, Z_{12}, and Z_{21}, where

$$Z_{mn} = \left.\frac{F_m}{\dot u_n}\right|_{\dot u_m = 0}, \qquad m, n = 1, 2.$$

The impedances Z are complex, frequency-dependent quantities. In sinusoidal vibration, the TEF at the endpoints 1 and 2 read

$$\begin{aligned} P_1 &= \tfrac{1}{2}\,\mathrm{Re}\{|\dot u_1|^2 Z_{11} + \dot u_1^* \dot u_2 Z_{12}\}, \\ P_2 &= \tfrac{1}{2}\,\mathrm{Re}\{|\dot u_2|^2 Z_{22} + \dot u_1 \dot u_2^* Z_{21}\}. \end{aligned} \tag{12}$$

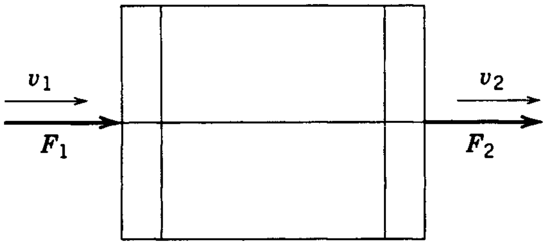


Fig. 6 Energy flow through a resilient element.

Here $|\cdot|$ and the asterisks denote absolute and complex-conjugate values, respectively, while Re denotes the real part of a complex quantity. If the motion of the endpoints is random stationary, the TEF spectra at the two points can be obtained simply by replacing the velocity products with the appropriate power and cross-spectral density functions and suppressing the common factor $\frac{1}{2}$.

At low frequencies, much below the first resonance of the element where it behaves essentially like a massless spring of stiffness K, the following formula applies for the TEF[13]:

$$P = K(u_1 - u_2)\dot u_1. \tag{12a}$$

The average net value of the TEF in this case is

$$\langle P \rangle = K\langle u_1 \dot u_2 \rangle. \tag{12b}$$

Equations (12a) and (12b) apply for an element with negligible damping. If damping is significant, a correction has to be added to the existing TEF expression (12a). For viscous damping, defined by a damping constant C, the correction reads

$$dP = C(\dot u_1 - \dot u_2)\dot u_2. \tag{12c}$$

A resilient element is often subjected to motion of a more general nature than the simple translatory motion assumed so far. Both forces and moments are generated at the endpoints and both contribute to the TEF. Moreover, coupling of motions usually takes place, which means that excitation in one direction produces additional movements in other directions as well. In such circumstances, evaluation of energy flowing through the element becomes increasingly difficult. The basic guideline for evaluation of the TEF in such a case is the following:

- Determine the appropriate impedance matrix (i.e., relationships between all the relevant forces + moments and translational + angular velocities).
- For each particular case, reconstruct the resultant force + moment vectors at the endpoints from the known motions of these points.
- The TEF is the scalar product of the force vector and the translational velocity vector plus the product of the moment vector and the angular velocity vector. This applies to each of the end points separately.

If the resilient element is highly nonlinear in the operating range, or if the provisions for attachment cannot be considered as points (e.g., rubber blocks of extended length, sandwiched between steel plates), the preceding approach becomes invalid. Some resilient elements (e.g., ones made of elastomers) have mechanical properties that are highly dependent not only on frequency but also on static preloading, temperature, and dynamic strain. All of these parameters should be taken into consideration when determining the impedances of the element.

4 MEASUREMENT PROCEDURES

Formulas for SI or EF [Eqs. (2)–(6) and (8)–(10)] contain constants (such as $E, \nu, h, \ldots$) and time-varying quantities that have to be measured (velocities, displacements, strains). Time-varying quantities are mutually linked as products. In order to obtain instantaneous values of SI or EF, all the varying quantities must be measured simultaneously. Net values of SI or EF can be obtained by evaluating the average value of each constituent product separately (where more than one product appears in the governing equation) and adding partial results off-line. Spectral representation of SI or EF can be achieved in a similar way using spectral densities instead of product averages.

The main practical problem is measurement of quantities that are spatial derivatives of velocities, displacements, and strains.

In order to measure these quantities, finite-difference approximations of spatial derivatives can be employed. Finite-difference approximations cause errors that can be analytically evaluated and sometimes kept within acceptable limits by an appropriate choice of parameters influencing the measurement accuracy.

Another way of measuring intensity/energy flow consists of establishing a wave model of the structure analyzed and fitting it with measured data. This inverse technique is known as the wave decomposition approach. In such a way all of the variables entering the intensity formulas can be identified.

4.1 Transducers for Measurement of SI (EF)

The quantities that need to be measured for evaluation of SI or EF, velocities (displacements) or strains, can be detected, in principle, in a variety of ways. In this respect, two basic approaches are possible: (a) use of transducers physically mounted on the object to be investigated and (b) detection of movement of the object by noncontact means.

Structural vibrations at a particular point in an object generally consist of both normal and in-plane (tangential) components, which can be translational and/or rotational. Measurement of SI or EF requires separate detection of some of these component motions, without any perceptible effect from other motions. This requirement poses the major problem in practical measurements. Another major problem results from the nonideal behavior of individual transducers such as phase shift between the physical quantity measured and the electrical signal that represents it. Both effects can severely degrade accuracy, since the measured quantities make up the product(s) where these effects matter more than in any straightforward measurement of the constituent quantities.

Seismic Accelerometer An accelerometer is easy to mount and to use. However, when using accelerometers for SI (FE) work, it is important to consider the following:

- Although the cross sensitivity of an accelerometer is usually low (up to a few percent), its effect can still be important in cases where one component of motion is measured in the presence of a strong component of cross motion (e.g., measurement of in-plane motion).
- Due to its finite size, the sensitivity axis of an accelerometer can never be positioned at the measurement surface but will always be somewhere above it. In this way, any in-plane rotation of the surface affects the output of the accelerometer. Compensation for this drawback can be effected by introducing a correction signal based on the measured rotation of the surface.
- An accelerometer dynamically loads the object to which it is attached. This effect is of importance for SI measurement on lightweight structures.

Strain Gauge Strain gauges measure normal strain elongation. Conventional (resistive) strain gauges exhibit few of the drawbacks associated with accelerometers: cross sensitivity is low and the sensitivity axis is virtually at the surface of the object, due to vanishingly small thickness of the gauge, while dynamic loading is practically zero. However, the conventional gauges exhibit low sensitivity where typical levels of strain induced by structural vibration are concerned, which means that typical signal-to-noise values could be low. Semiconductor gauges have a much higher sensitivity than the conventional type, typically two orders of magnitude higher. High dispersion of sensitivity makes the semiconductor gauges unacceptable for intensity work.

Shear strain at the free surface can be measured by two strain gauges positioned at 45° and 135° with respect to the direction to be considered. The shear strain is then equal to the difference between the two normal strains measured by the gauges. The difference signal can be obtained directly from the gauge conditioner by wiring the two gauges in adjacent branches of the gauge bridge.

Noncontact Transducers A range of various contactless transducers are available for the detection of structural motion such as inductive, capacitive, eddy current, light intensity, and light-modulated transducers. The majority of contactless transducers are of a highly nonlinear type, where linearity could be achieved by compensating electronic circuitry.

Optical transducers based on the interference of coherent light such as LDV (laser Doppler velocimeter) devices are well suited for SI work. The most promising are however optical techniques for whole-field vibration detection, such as holographic or speckle-pattern interferometric methods.[14] Applications of these are limited for the time being to plane surfaces only.

Another interesting approach for noncontact whole-field SI measurement consists in using acoustical holography.[15,16] This is again possible when the structure analyzed is of a simple regular shape such as a plate or cylinder.

4.2 Transducer Configurations

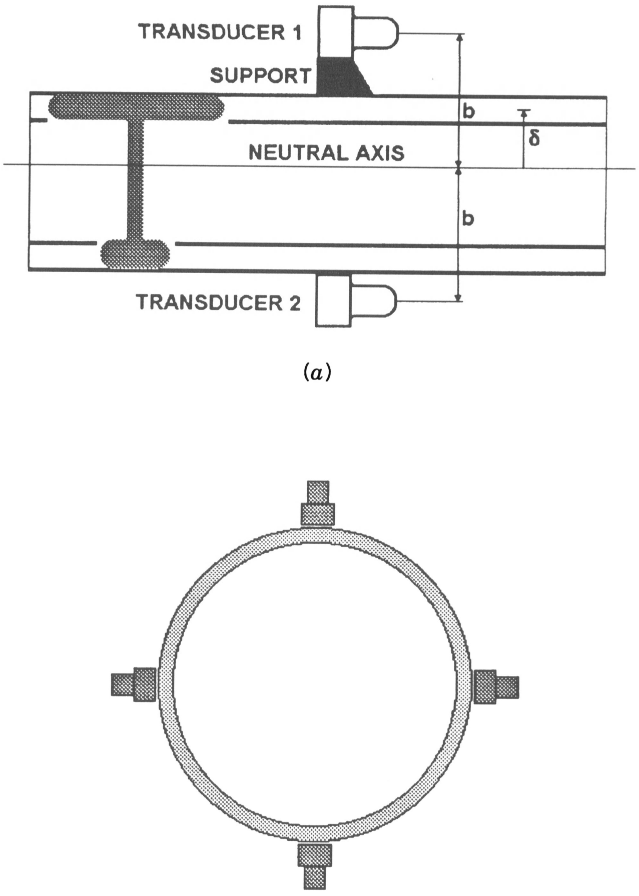
In the measurement of SI (EF), three situations arise where more than one transducer is needed for measurement of a single physical quantity:

- When separation of a particular type of motion from the total motion is required
- When detection of a spatial derivative is required
- When different wave components need to be extracted from the total wave motion

The first case arises with beams, plates, and shells where longitudinal and flexural motions take place simultaneously (in the case of shells, this is always the case) and where corrections of the type of Eq. (7) are not appropriate. The second case occurs when a direct measurement implementation of SI (EF) governing equations is performed. The third case applies to waveguides (beams, pipes etc.) where the wave decomposition approach is used for EF evaluation.

Extraction of a Single Type of Motion It has been shown in Section 3.6 how the longitudinal component of either motion or strain, denoted by 0, can be separated from the total motion measured at the surface.

An elegant way to circumvent the difficulty of measuring the correcting term consists in placing two identical transducers at opposite sides of the neutral plane and adding or subtracting the outputs from the transducers. In this way longitudinal and flexural strains can be separated (Fig. 7*a*). Figure 7*b* shows a similar scheme for separating pulsation-induced from flexural-induced motions of a vibrating pipe.



(*a*)

(*b*)

Fig. 7 Transducer configuration for extraction of a simple type of motion.

Detection of Spatial Derivatives Formulas for SI(EF) contain spatial derivatives that cannot be measured directly. The usual way of measuring spatial derivatives is by using finite-difference concepts.

The spatial derivative of a function q in a certain direction (e.g., the x-direction) can be approximated by the difference between the values of q at two adjacent points, 1 and 2, located on a straight line in the direction concerned:

$$\frac{\partial q}{\partial x} = \frac{q_2 - q_1}{d}, \qquad d \ll \text{wavelength}. \tag{13}$$

The spacing d should be small in order to make the approximation (13) valid. The term "small," however, is not related to any absolute value, but rather to a value relative to the wavelength of the structural motion. The larger the wavelength (i.e., the lower the frequency), the better the approximation (13).

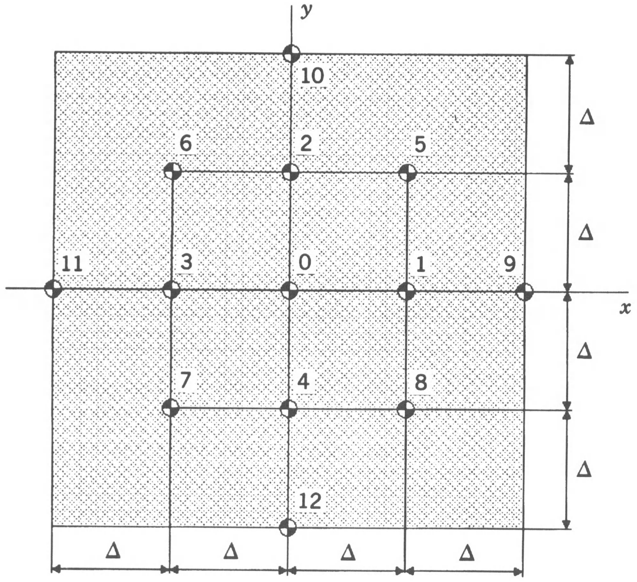


Fig. 8 Transducer configuration for measurement of spatial derivatives of vibration.

Using the principle applied in Eq. (13), higher order derivatives can be determined from suitable finite-difference approximations. Referring to the scheme shown in Fig. 8, the spatial derivatives appearing in the EF formulas can be determined from the finite-difference approximations listed below:

$$\begin{aligned}
\frac{\partial q}{\partial x} &\approx \frac{1}{2\Delta}(q_1 - q_3),\\
\frac{\partial q}{\partial y} &\approx \frac{1}{2\Delta}(q_2 - q_4),\\
\frac{\partial^2 q}{\partial x^2} &\approx \frac{1}{\Delta^2}(q_1 - 2q_0 + q_3),\\
\frac{\partial^2 q}{\partial y^2} &\approx \frac{1}{\Delta^2}(q_2 - 2q_0 + q_4),\\
\frac{\partial^2 q}{\partial x\,\partial y} &\approx \frac{1}{4\Delta^2}(q_5 - q_6 + q_7 - q_8), \qquad (13a)\\
\frac{\partial(\Delta q)}{\partial x} &\approx \frac{1}{2\Delta^3}(q_5 + q_8 + q_9 - 4q_1 - q_6\\
&\quad - q_7 - q_{11} + 4q_3),\\
\frac{\partial(\Delta q)}{\partial y} &\approx \frac{1}{2\Delta^3}(q_5 + q_6 + q_{10} - 4q_2 - q_7\\
&\quad - q_8 - q_{12} + 4q_4).
\end{aligned}$$

Measurement of higher order derivatives, such as those appearing in the expressions for plate and shell flexural vibration, becomes difficult in practice due to the large number of transducers required. For example, to measure the Laplacian derivative, eight transducers are needed. In addition, measurement errors increase exponentially with the order of derivative. For these reasons, measurement of higher order spatial derivatives should be avoided. Simplified measurement techniques, such as the two-transducer method described in Section 3.7, may seem far more suitable for practical purposes. It should be pointed out, however, that simplifications assume far-field conditions, that is, locations far from sources or discontinuities. Very often the near-field regions will be of prime interest from the point of view of the applicability of results. Therefore simplified procedures should be used with care.

Wave Decomposition Technique This approach is suitable for waveguides (rods, beams, and pipes) where the vibration field can be described by a simple wave model in the frequency domain.[17]

Nondispersive vibration, longitudinal and torsional, will then be represented by two oppositely propagating waves:

$$\begin{aligned}
\xi(x,\ \omega) &= [A_+ \exp(-jkx + \varphi_+) + A_- \exp(jkx + \varphi_-)]\\
&\quad \cdot \exp(j\omega t), \qquad j = \sqrt{-1}. \qquad (14)
\end{aligned}$$

In such a representation the amplitudes A and phases φ of the two waves are unknown but the wavenumber k is supposed to be known: $k = \omega/c$, where c is the wave velocity. For longitudinal vibration $c = \sqrt{E/\rho}$, while for torsional vibration $c = \sqrt{G/\rho}$, where G is the shear modulus.

When flexural vibrations are concerned, the model (14) is valid only in the far field. If the far-field conditions are not met, it should be appended by two additional terms representing evanescent waves that decay from the two boundaries, $D_+ \exp(-kx + \phi_+)$ and $D_- \exp(kx + \phi_-)$.

Once the model of vibratory motion has been established, it can be substituted in the appropriate EF governing formula (Section 3). The EF of nondispersive vibration then becomes proportional to the difference of squares of two amplitues:

$$P_x = \rho c S \omega^2 (A_+^2 - A_-^2). \qquad (15)$$

The EF of flexural waves will be contributed additionally by the evanescent terms:

$$P_{x,\text{flexural}} = JEk^3\omega[A_+^2 - A_-^2 - D_+D_- \sin(\phi_+ - \phi_-)], \qquad (15a)$$

where $k = (\rho S/JE)^{1/4}\sqrt{\omega}$ is the flexural wavenumber. All that needs to be done by measurement is to determine the unknown amplitudes A_+ and A_- and, if needed, the evanescent components D_+, D_- and $\phi_+ - \phi_-$.

If two transducers are placed at two arbitrary points 1 and 2, the amplitude square difference will be proportional to the imaginary part of their product $u_1 u_2^*$, which corresponds to the cross spectral density between the two signals:

$$A_+^2 - A_-^2 = \frac{\mathrm{Im}\{u_1 u_2^*\}}{\sin(\omega d/c)}, \tag{16}$$

where the asterisk indicates the complex-conjugate form and d is the distance between the two points. It should be noted that Eqs. (15) and (16) represent an exact measure of EF in a nondispersive vibration field.

To measure flexural EF in the presence of a nonnegligible evanescent field, four instead of two transducers will be needed at different locations to determine all four wave amplitudes as well as phase difference $\phi_+ - \phi_-$. The evanescent vibration components will not be negligible in the frequency range where the wavelengths of vibration are larger than approximately one-half of the length of the measured structure.

4.3 Measurement Errors

Instrumentation Errors In making SI (EF) measurements, various types of errors may occur. In the preceding section, three different sources of inaccuracy were outlined: sensitivity to cross motions, off-set of the sensitivity axis due to finite transducer size, and dynamic loading due to finite transducer mass.

The measurement inaccuracies caused by transducer imperfections are not the only ones related to measurement equipment. An electrical signal from the transducer may be subjected to distortion and noise, which can greatly affect the resulting intensity readings.

The most significant source of signal distortion where SI (EF) is concerned is the phase shift between the physical quantity measured and the conditioned electrical signal that represents it. Since SI (EF) evaluation always required multiplication of signals, the phase shift will directly affect the result, altering the true physical phase angle between the original quantities. This effect becomes particularly significant in connection with finite-difference representation of some of the quantities.

The relative error in measuring the first spatial derivative due to the instrumentation phase shift $\Delta\varphi$ is

$$\varepsilon_{\Delta\varphi} = \frac{\Delta\varphi}{kd\cos\alpha}, \tag{17}$$

where k is the wavenumber, d the distance between transducers, and α the angle between the direction of the derivative and the wave direction.

The error is linearly proportional to the instrumentation phase shift and inversely proportional to the product of the wavenumber and the transducer spacing. The error due to phase mismatch in measurement of higher order derivatives will be of a form similar to Eq. (19), providing $\Delta\varphi$ remains small:

$$\varepsilon_{\Delta\varphi} = \frac{\Delta\varphi}{(kd\cos\alpha)^n}, \qquad n = \text{order of derivative.} \tag{17a}$$

To keep the phase mismatch error small, the instrumentation phase shift between channels must be extremely small where measurement of higher order derivatives is concerned. This is because the product kd, which appears in the denominator of Eq. (19a), has itself to be small in order to make the finite-difference approximation valid. For example, a two-transducer method (i.e., $n = 1$) was found to be appropriate when the phase shift was of the order of 0.1°.[18]

Another important error that can affect measurement accuracy can be caused by the mass loading excited by the measuring transducer.[19] The transducer mass should be kept well under the value obtained by dividing the impedance of the structure by the radian frequency.

Finite-Difference Convolution Error By making approximations of spatial derivatives using finite differences, systematic errors are necessarily produced that depend on the spacing between the transducers and the nature of the vibratory field concerned. For example, the first spatial derivative of the x-direction, approximated by finite difference according to Eq. (18), will contain an error equal to

$$\varepsilon_{\partial u/\partial x} \approx \frac{d^2}{24}\,\frac{\partial^3 u/\partial x^3}{\partial u/\partial x}. \tag{18}$$

In plane-wave motion, with the wave direction occurring at an angle α to the axis, the error reads approximately

$$\varepsilon_{\partial u/\partial x} \approx -\tfrac{1}{24}k^2 d^2 \cos^2\alpha, \qquad \varepsilon \ll 1. \tag{18a}$$

For flexural vibration, the wavenumber is proportional to the square root of the frequency. The error in measuring the first spatial derivative will therefore be proportional to the product of the frequency and the square of the transducer spacing, d. It depends on the angle of wave propagation α. The error is seen always to be negative; that is, the finite-difference technique underestimates the true value.

The finite-difference error for the second derivative is approximately twice the error for the first derivative:

$$\varepsilon_{\partial^2 u/\partial x^2} \approx 2\varepsilon_{\partial u/\partial x} \approx -\tfrac{1}{12}k^2 d^2 \cos^2 \alpha. \qquad (18b)$$

The errors of higher order derivatives rise in proportion to the derivative order, in contrast to instrumentation errors, which exhibit an exponential rise.

The finite-difference error for any spatial derivative can be shown to be proportional to the square of the transducer spacing d. In addition, where plane waves are concerned, the error is proportional to the square of the wavenumber.

Wave Decomposition Coincidence Error The technique of wave decomposition is essentially a solution to an inverse problem. This solution can always be represented as a ratio of two frequency-dependent functions. The problem arises at frequencies where the denominator approaches zero, that is, where the solution is badly conditioned. At these particular coincidence frequencies, which correspond to a specific relationship between the geometry of the transducer array used and the wavelength, even very small measurement imperfections can cause significant measurement error.

For the simple case of nondispersive vibration motion [Eq. (14)], the frequency singularities occur for $\omega d/c = n\pi$, that is, each time half the wavelength becomes an integer fraction of the spacing between the two transducers. More complex transducer arrays will result in larger number of coincidence frequencies that need not be equidistant. The larger are the distance(s) between the transducers, the more numerous are the coincidence frequencies within a given frequency band.

Total Measurement Error For small error values due to instrumentation imperfections and measurement method imperfections, finite-difference convolution or wave decomposition coincidence, the total error will be the simple sum of the two errors. The total error is therefore subject to the same spacing/wavenumber dependence as errors in the measurement of individual quantities.

REFERENCES

1. S. P. Timoshenko and J. N. Goodier, Theory of Elasticity, Int. Student Ed., McGraw-Hill, New York, 1970.
2. D. Noiseux, "Measurement of Power Flow in Uniform Beams and Plates," *J. Acoust. Soc. Am.*, Vol. 47, 1970, pp. 238–247.
3. G. Pavic, "Determination of Sound Power in Structures: Principles and Problems of Realisation," *Proc. 1st Congress on Intensity Techniques*, Senlis (France), September–October 1981, pp. 209–215.
4. G. Pavic, "Measurement of Structure-Borne Wave Intensity, Part 1: Formulation of Methods," *J. Sound Vib.*, 1976, Vol. 49, pp. 221–230.
5. P. Rasmussen and G. Rasmussen, "Intensity Measurements in Structures," *Proc. 11th Int. Congress on Acoustics*, Paris, July 1983, Vol. 6, pp. 231–234.
6. C. R. Fuller and F. J. Fahy, "Characteristics of Wave Propagation and Energy Distribution in Cylindrical Elastic Shells Filled with Fluid," *J. Sound Vib.*, 1982, Vol. 81, pp. 501–518.
7. G. Pavic, "Vibrational Energy Flow in Elastic Circular Cylindrical Shells," *J. Sound Vib.*, 1990, Vol. 142, pp. 293–310.
8. G. Pavic, "Measurement of Structure-Borne Acoustical Intensity in Thin Shells," *Proc. of Inter-Noise 92*, Toronto, July 1992, Vol. 1, pp. 509–514.
9. I. Levanat, "Vibrational Energy Flow in Spherical Shell," *Proc. 3rd Congress on Intensity Techniques*, Senlis (France), August 1990, pp. 83–90.
10. J. W. Verheij, "Cross-Spectral Density Methods for Measuring Structure-Borne Power Flow on Beams and Pipes," *J. Sound Vib.*, 1980, Vol. 70, pp. 133–139.
11. C. A. F. de Jong and J. W. Verheij, "Measurement of Energy Flow Along Pipes," *Proc. Int. Congress on Recent Develop. in Air and Structure-Borne Sound and Vibration*, Auburn (USA), 1992, pp. 577–584.
12. R. J. Pinnington and R. G. White, "Power Flow through Machine Isolators to Resonant and Non-Resonant Beams," *J. Sound Vib.*, 1981, Vol. 72, pp. 179–197.
13. G. Pavic and G. Oreskovic, "Energy Flow through Elastic Mountings," *Proc. 9th Int. Congress on Acoustics*, Madrid, 4–9 June 1977, Vol. 1, p. 293.
14. J. C. Pascal, X. Carniel, V. Chalvidan, and P. Smigielski, "Energy Flow Measurements in High Standing Vibration Fields by Holographic Interferometry," *Proc. Inter-Noise 95*, Newport Beach, July 1995, Vol. 1, pp. 625–630..
15. E. G. Williams, H. D. Dardy, and R. G. Finck, "A Technique for Measurement of Structure-Borne Intensity in Plates," *J. Acoust. Soc. Am.*, 1985, Vol. 78, pp. 2061–2068.
16. J. C. Pascal, T. Loyau, and J. A. Mann III, "Structural Intensity from Spatial Fourier Transform and Bahim Acoustic Holography Method," *Proc. 3rd Int. Congress on Intensity Techniques*, Senlis (France), August 1990, pp. 197–206.
17. C. R. Halkyard and B. R. Mace, "A Wave Component Approach to Structural Intensity in Beams," *Proc. 4th Int. Congress on Intensity Techniques*, Senlis (France), August–September, 1993, pp. 183–190.
18. G. P. Carroll, "Phase Accuracy Considerations for Implementing Intensity Methods in Structures," *Proc. 3rd Congress on Intensity Techniques*, Senlis (France), August 1990, pp. 241–248.
19. R. J. Bernhard and J. D. Michol, "Probe Mass Effects on Power Transmission in Lightweight Beams and Plates," *Proc. 3rd Congress on Intensity Techniques*, Senlis (France), August 1990, pp. 307–314.

75

ACTIVE VIBRATION CONTROL

C. R. FULLER

1 INTRODUCTION

Passive control of vibrations in structures including isolation mounts have been discussed in Chapter 71 and Ref. 1. These methods work well at high frequencies or in a narrow frequency range but often have the disadvantage of added weight and poor low-frequency performance. Active vibration control has demonstrated the potential to solve many of these problems. Although the potential of active vibration control has been well known for many years, recent advances in fast digital signal processing (DSP) computer chips have made these systems realizable.

In active vibration control secondary vibration inputs are applied to the structure in order to modify its response in a desired manner. The major components of an active vibration control system are the plant, actuators, sensors, and a controller. The plant represents the physical system to be controlled. Error sensors are needed to measure the system response while control actuators provide the necessary inputs to the plant to modify its response. The controller implements the chosen control algorithm to ensure that the controlled (or closed-loop) system behaves as required. The closed-loop system consists of the open-loop (uncontrolled) system dynamics combined with the dynamics of the controller and thus behaves in a modified form. These components will be discussed separately below and then unified into a coupled system behavior by use of example applications. Basic theory for each of the components is also discussed in order to provide the framework for analyzing controlled system behavior.

The arrangement of an active control system is usually based upon the physics of the system to be controlled and thus, as illustrated below, is often application dependent. There are many texts on active control of vibration, mainly focused on the control theory. The text by Meirovitch[2] provides an excellent theoretical basis while the text of Fuller et al.[3] describes in detail both feedforward and feedback control theory, actuators and sensors as well as many relevant applications. Chapter 84 also discusses the related area of active noise control. It should be noted that active control is not a panacea for every vibration problem, and its application should be chosen with care depending upon a number of important characteristics such as the temporal nature of the disturbance, the spatial nature of the noise field, and the dynamic characteristic of the disturbance.[4]

2 CONTROL ACTUATORS

Actuators are used to introduce control forces into the plant in order to modify its behavior, and their design/selection (in conjunction with the sensors) is often the most important step in an active noise control project. Actuators can take various forms that are dependent upon system requirements such as the required control authority (amount of control force, moment, strain, or displacement), power, frequency response, and physical constraints such as size, mounting requirements, and so on. Actuators are generally classified into two main categories: *fully active* actuators that apply a secondary vibrational response to the structure and *semiactive* actuators that are passive elements and can be used to adaptively adjust the mechanical properties of the system.

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker
Copyright © 1997 John Wiley & Sons, Inc.

2.1 Electrodynamic Actuators

Electrodynamic actuators or shakers consist of a moving wire coil mounted inside a permanent magnet. As they are readily available and their behavior is relatively well understood (see Ref. 5 and Chapters 63 and 72), electrodynamic shakers are presently the most common control input as a fully active device. They are usually installed by either attaching directly to the structure by a stinger or placing in series or parallel with passive mounts, as in the case of active isolation (see Section 5.1). While electrodynamic actuators have the advantage of relatively large control displacement, they are usually bulky and require some form of support structure. However, there are many applications at low frequencies in which their use is more than adequate. Electrodynamic actuators are usually modeled in terms of control inputs as point forces and they are thus "spectrally white" in terms of the spatial wavenumber response of the excited structure.

Typical structural components encountered in practice can often be simply represented as beams or plates. The transverse response w of a simply supported thin beam to an oscillating point force $Fe^{i\omega x}$, representative of a control shaker, can be described as[6]

$$w(x,t) = \frac{-2F}{M}\sum_{n=1}^{\infty}\frac{\sin(n\pi x_f/L)\sin(n\pi x/L)}{\omega^2-\omega_n^2+i2\zeta_n\omega_n\omega}e^{i\omega t}, \quad (1)$$

where x is the axial coordinate, x_f is the location of the force, L is the length of the beam, ω is the frequency of excitation, M is the total mass of the beam, and ζ_n is the modal damping ratio. In Eq. (1), ω_n is the natural frequency of the nth mode of the beam.[6] The transverse vibration displacement w of a simply supported thin panel excited by a point force $Fe^{i\omega t}$ is also described by[6]

$$w(x,y,t) = \frac{-4F}{M}\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \cdot \frac{\sin(m\pi x_f/L_x)\sin(n\pi y_f/L_y)\sin(m\pi x/L_x)\sin(n\pi y/L_y)e^{i\omega t}}{\omega^2-\omega_{mn}^2+i2\zeta_{mn}\omega_{mn}\omega}, \quad (2)$$

where the force is located at (x_f, y_f), ω_{mn} is the natural frequency of the (m,n)th mode, and M is the total mass of the plate. Response of other more complex systems to point forces can be found in well-known texts such as Ref. 6.

2.2 Piezoelectric Actuators

As discussed in Ref. 7, piezoelectric transducers consist of material that expands or contracts when an electric field is applied over it. By applying an oscillating voltage to the piezoelectric element, its strain can be made to oscillate at the same frequency as the input. Three major types of piezoelectric material are readily available: (1) a ceramic form such as PZT, which has relatively high control strain but is brittle; (2) a polyvinyl form such a PVDF, which is flexible but has less control strain for the same configuration; and (3) a piezoelectric rubber, which is useful for underwater applications. Piezoelectric material is configured in the two main forms of *stack* and *wafer*, as shown in Fig. 1. Piezoelectric stacks are configured so that when a voltage is applied across the electrodes, the stack usefully expands in its long or 3–3 axis.[7] Stack arrangements such as Fig. 1 are thus suitable for actuators in vibration isolation in the two configurations of parallel and series shown in Figs. 1*a* and 1*b*.

For a given applied voltage V the net static displacement δ of the piezoelectric ceramic actuator shown in the parallel configuration of Fig. 1*a* can be calculated from

$$\delta = \frac{d_{33}V + F/K_a}{1 + K/K_a}, \quad (3)$$

where d_{33} is the strain constant of the piezoelectric material in the 3–3 axis,[7] K is the external spring stiffness, K_a is the actuator stiffness ($K_a = E_aA_a/L_a$, where L_a, A_a, and E_a are the actuator length, cross sectional area, and Young's modulus, respectively); and F is the external load force. Note Eq. (3) can be used as the basis of an approximate dynamic analysis, as discussed in Ref. 3. The significant advantage of the piezoelectric stack is that it can provide high force; however, its displacement is limited (relative to electrodynamic actuators), which, as discussed later, implies that its use is primarily in series active isolation implementations.

The other common form of piezoelectric actuator is the wafer arrangement shown in Fig. 1*c*. In this form the transducer usefully strains in its thin transverse axis (3–1 or 3–2) when a voltage is applied across the electrodes. The wafer is usually bonded or embedded directly into the structure and actuates the structure by applying surface or interior strains. Figure 1*c* shows piezoelectric elements arranged to create pure bending. In this configuration two colocated elements are employed and wired 180° out of phase so that as one element expands, the other contracts. As discussed in Ref. 8, the surface-mounted colocated actuator for a perfect bonding layer (which is a good approximation if the glue layer is thin) effectively applies a line moment to the structure at the boundaries of the actuator. The magnitude of this linear

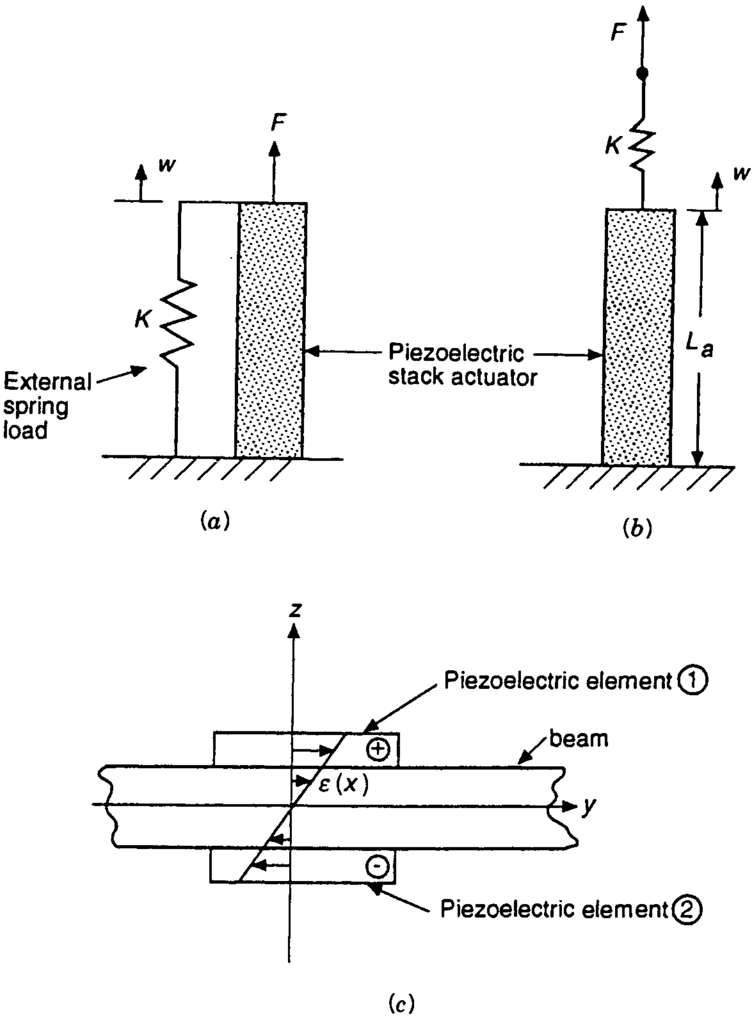

F
w
K
External spring load
Piezoelectric stack actuator
L_a
(a)
(b)
z
Piezoelectric element ①
beam
⊕
ε (x)
y
⊖
Piezoelectric element ②
(c)

Fig. 1 Piezoelectric actuator configurations: (*a*) stack in parallel; (*b*) stack in series; (*c*) colocated wafer.

moment density for a one-dimensional structure such as a simply supported beam is given by[9]

$$m_x = \frac{-E_b 2Ph^2\epsilon_{pe}}{3(1-P)}, \tag{4}$$

where the piezoelectric element unconstrained strain $\epsilon_{pe} = d_{31}V/t$. The constant P is defined as

$$P = \frac{-3th(2h+t)}{2(h^3+t^3)+3ht^2}\,\frac{E_{pe}}{E_b}. \tag{5}$$

In Eqs. (4) and (5) E_b and E_{pe} are the Young's elastic modulus of the beam material and piezoelectric element, respectively, while h is the half thickness of the beam, t is thickness of the piezoelectric element, V is the applied voltage, and d_{31} is the piezoelectric transverse strain constant. For a two-dimensional patch system in spherical bending of a plate the effective moments m_x, m_y of the actuator are given by[10]

$$m_x = m_y = \frac{-E_p(1+\nu_{pe})}{3(1-\nu_p)}\,\frac{2P'h^2\epsilon_{pe}}{(1+\nu_p-(1+\nu_{pe})P')}, \tag{6}$$

where

$$P' = P\,\frac{1-\nu_p^2}{1-\nu_{pe}^2}. \tag{7}$$

In Eqs. (6) and (7) ν_{pe} and ν_p are the Poisson's ratio of the piezoelectric and the plate material, respectively. More information of piezoelectric elements and their properties as well as definitions of important terms can be found in Ref. 7. Excitation of beam and plate structures with piezoelectric patches are discussed in Refs. 9 and 10. Piezoelectric actuators can be used either in a semiactive or a fully active form.

2.3 Advanced Actuators

The desire for an optimum actuator has stimulated development of new materials and configurations. The requirements, depending upon application, are typically high force and/or displacement, large bandwidth of operation, lower power consumption, and ease of use.

Shape memory alloy (SMA) shows much potential for high-force actuation at DC to very low frequencies. The mechanism by which SMA fibers or films exhibit their characteristic shape memory effect can be described very basically as follows: An object in the low-temperature martensitic condition, when plastically deformed and the external stresses removed, will regain its original (memory) shape when heated. The process is the result of a martensitic phase transformation taking place during heating. When SMA is heated through its transformation temperature, the elastic modulus of the material changes dramatically, and if the material is initially stretched, it will shrink back to its original size. If the SMA is constrained during this procedure, very high restoring forces are generated. More information on SMA can be found in Ref. 11.

The SMA systems are usually configured by embedding the material in a structue in a similar manner to layups of composite structures. It can be activated in a steady-state manner, applying steady-state in-plane forces that can be used to tune the resonant frequencies or change the mode shapes of the system (i.e., as semiactive actuator.[12] Over a very limited frequency range it can be used as an oscillating control input (i.e., a fully active actuator) by positioning the SMA off the central axis of the structure and driving with an oscillating voltage.[12] The advantage of the SMA is the extremely high force and deflection that it can provide at very low frequencies. Disadvantages are high power consumption, a need for heat dissipation, and a limited frequency response, which depends entirely upon the cooling rate. Use of SMA in the control of transient vibration of a cantilevered beam as well as sound radiation control is discussed in Ref. 12.

Magnetostrictive materials also show much possibility as advanced fully active actuators in that they fill the performance gap between low-force, high-displacement electromagnetic actuators and high-force, low-displacement piezoelectric devices. A typical actuator using the material Terfenol-D and approximately 11 cm length and 5 cm diameter can produce forces in excess of 450 N over a frequency range of 0–2.5 kHz.[13] However, the power requirements tend to be somewhat higher than electrodynamic and piezoelectric devices.

Electrorheological (ER) fluids are suspensions of highly polarized fine particles dispersed in an insulating oil.[14] The viscosity and the elasticity of the ER fluid can be changed several orders of magnitudes when an electric field is applied to the medium. The ER fluid can thus be embedded in structural systems and used as a "semiactive" actuator to tune the system's properties such as damping and stiffness by varying the voltage applied to the ER fluid.[15]

3 ERROR SENSORS

Error sensors are employed to measure the motion of the system to be controlled. This information either is used directly as the variable(s) to be minimized or is used to calculate a related state of the system to be controlled. Sensors come in three main forms: point sensors, arrays of point sensors, and distributed sensors. Choice of the particular configuration is dependent upon the system variable to be controlled, as discussed below.

3.1 Accelerometers, Force Transducers, and Impedance Heads

Accelerometers are commonly used as error sensors due to the ease of use and reliabilty of performance. Their implementation and behavior is discussed in more detail in Chapter 72. When attached to a vibrating structure, they can provide estimates of the time-varying acceleration (as well as velocity and displacement with signal processing) at the point of attachment. Due to their small size, at low frequencies accelerometers provide a point sampling of the structural motion and thus are "spectrally white" in wavenumber response. Thus an individual accelerometer will equally sense all structural wavenumber or modally weighted contributions. Use of an accelerometer output as an error signal will result in direct control of the structural motion at that point.

Force transducers can also be used as error sensors, particularly in applications such as active isolation, where they may be placed in series with passive isolation. Minimization of the force transducer output will result in zero dynamic force being applied to the receiving structure. The point input impedance of a structure such as at a disturbance location can be estimated using an impedance head that consists of a colocated accelerometer and force transducer. The output of the impedance head can then be used as a narrow-band error sensor variable in order to minimize the real part of the structural input impedance and thus disturbance input power flow into the structure.

Choice of one of the above transducers as an error sensor is dependent upon the form of application. This will be illustrated in the later example applications.

3.2 Arrays of Point Sensors

Usually it is more effective to use accelerometers as error sensors in an array configuration whose output signal is processed to provide some global or distributed state associated with the system.

The *kinetic energy* of a thin vibrating system can be minimized by using a distributed array of accelerometers positioned over the required controlled domain. An estimate proportional to out-of-plane kinetic energy of the structure is given by

$$E = \sum_{i=1}^{N} |\dot{w}_i|^2, \tag{8}$$

where N is the total number of accelerometers and $\dot{w}$ is the out-of-plane velocity obtained by integrating the acceleration signal. The variable E can be minimized, or if an N-channel controller is available, each accelerometer can be used as an individual control channel. The required number and spacing of accelerometers are determined by the Nyquist sampling criterion applied in the spatial domain: two accelerometers must be spaced within at least a half-wavelength of the highest mode required to be observed and controlled. Alternatively, the accelerometers can be appropriately distributed throughout the surface whose global motion is to be controlled. Use of a lower number of accelerometers (relative to the number of actuators) often leads to large minimization at their locations but increased levels of vibration at other locations. This effect is termed control spillover. In general, to ensure controllability, a system state or mode has to be individually observed.

Wavenumber components of structural motion can be obtained using arrays of equispaced accelerometers arranged along a coordinate axis. The spatial one-dimensional Fourier transform can be approximated using a discrete Fourier transform such as[3]

$$\hat{W}(k) = \sum_{n=1}^{K} w_m(x_n) e^{ikx_n} \Delta x, \tag{9}$$

where $w_m(x_n)$ is the measured out-of-plane complex displacement at position x_n. In Eq. (9) there are N measurement points equally spaced at a distance Δx. Note Eq. (9) can also be expanded into a time–space transform if necessary. The outpt of Eq. (9) can be used to provide an error signal. Minimization of spectral components $W(k)$ then will provide control of selected wavenumber values. For example, in a finite structure vibrating at a single frequency, the wavenumber components will be discrete and associated with particular modes or waves traveling in the positive or negative directions. In many situations it is appropriate to control these components directly.

Modal decomposition of the structural motion is also a useful processing technique. To achieve this, an array of accelerometers is used to measure out-of-plane motion at a number of points. Thus, knowing the structural mode shapes and using a pseudoinverse technique, the modal amplitudes can be estimated from[16]

$$[A] = ([L]^T[L]^{-1})[L]^T[w_{me}], \tag{10}$$

where $[A]$ is a vector of m unknown modal amplitudes and $[w_{me}]$ is a vector of n measured displacements at (x_n, y_n). In Eq. (10) $[L]$ is an $n \times m$ matrix associated with the mode shapes Ψ of the system, defined as

$$[L] = \begin{bmatrix} \Psi_1(x_1,y_1) & \Psi_2(x_1,y_1) & \cdots & \Psi_m(x_1,y_1) \\ \vdots & & \ddots & \\ \Psi_1(x_n,y_n) & & & \end{bmatrix}. \tag{11}$$

Equation (10) can also be readily modified to decompose narrow-band traveling-wave fields into individual wave components and can be implemented for broadband random disturbances, as outlined in the work of Gibbs et al.[17] Equation (10) can thus provide indirect error information of a global nature associated with each mode. Minimization of a modal estimate will ensure that a particular mode is globally controlled in a vibrating structure. This is of use when particular modes are more important than others in a control strategy.

Kalman filtering techniques are used extensively in linear quadratic (LQ) control problems,[3] where all the internal state variables of the system to be controlled must be known. When all the states are not measurable, they have to be estimated using a mathematical model of the dynamic system and measurements of the output variables. For Gaussian disturbances, the Kalman filter is the optimal estimator that minimizes the variance of the estimation error.[18]

Power flow can be estimated from an array of accelerometers mounted on the structure. For the simplest case of traveling waves on a thin beam the net narrow-band bending power flow can be calculated by using two measurement points and finite-difference theory, as described in Chapter 74 and Ref. 19:

$$P(\omega) = \frac{2EIk_f^2}{\Delta x \omega^3} \operatorname{Im}(S_{21}). \tag{12}$$

In Eq. (12) k_f is the flexural wavenumber at frequency ω,[6] Δx is the spacing of the two acceleration measurement points, and S_{21} is the cross spectrum between the output of the two accelerometers. Use of Eq. (12) as an error signal will thus result in minimization of net power flow in the structure. Extension of Eq. (12) to include near fields is also discussed in Ref. 19 (see also Chapter 74). As discussed above, it is also possible to decompose the motion of the structure into wave component amplitudes using a pseudoinverse technique. The power flow associated with each direction of propagation or wave type can be then estimated and used as an error signal. Using approaches such as these, it is possible to actively control termination impedances of beams.[20]

3.3 Distributed Piezoelectric Sensors

Piezoelectric material, when strained from its free state, also creates a charge.[7] Thus, when configured properly, piezoelectric material with attached electrodes to collect the charge can be used as an error sensor. The most common arrangement is to bond the piezoelectric material to one side of the surface of the structure whose motion is to be sensed. If the piezoelectric material is very thin, then it can be assumed that its presence has no effect on the motion of the structure. For a two-dimensional structure covered with a finite layer of piezoelectric material the charge output $q(t)$ of the sensor (in rectangular coordinates) can be calculated from[21]

$$q(t) = (h_p + h_s) \iint_S \Gamma(x,y) \cdot \left[\epsilon_{31} \frac{\partial^2 w}{\partial x^2} + \epsilon_{32} \frac{\partial^2 w}{\partial y^2} + 2\epsilon_{36} \frac{\partial^2 w}{\partial x \partial y} \right] dx\, dy, \tag{13}$$

where $w(x,y)$ is the plate out-of-plane response, $\Gamma(x,y)$ is the sensor shape function, h_p and h_s are the thickness of the plate and sensor, and $\epsilon_{31}, \epsilon_{32}$, and ϵ_{36} are the piezoelectric stress–charge coefficients where the superscripts 1, 2, 3, refer to the z, x, and y directions of a three-dimensional coordinate system.[7] In effect, the piezoelectric material integrates the surface strain of the structure over its area S (i.e., in the 1–2 plane) and thus provides continuous distributed sensing. The actual voltage output of the piezoelectric material is dependent upon the electrical circuitry used to measure the piezoelectric output. Shaped piezoelectric sensors are surface-mounted sensors that are of specified finite size and shape. Particular motions or modal components of the vibrating structure can be observed by cutting the piezoelectric material into a specialized shape. The charge output of an example rectangular piezoelectric strip mounted on a simply supported plate so that it extends completely over the plate in the x-direction but is narrow in the y-direction is given by[22]

$$q(t) = K \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \frac{L_x L_y}{nm\pi^2} [\cos(m\pi) - 1] \cdot \left[\cos\left(\frac{n\pi d}{L_y}\right) - \cos\left(\frac{n\pi c}{L_y}\right) \right], \tag{14}$$

where L_x, L_y is the dimension of the plate in the x, y directions, K is a constant that modifies the charge output due to the external electrical load impedance, m and n are modal orders, and d and c are the vertical locations of the sensor boundaries in the y-direction. Note Eq. (14) shows that for all even modes in the x-direction ($m = 2$, 4, 6, ...) the sensor output is zero. Thus the sensor acts as a spatial wavenumber or modal filter only observing odd modes in the x-direction. Positioning and sizing of the piezoelectric strip can be adjusted to observe required modal or wavenumber components.

In order to observe individual particular modes of motion, the piezoelectric strip can be cut into a shape corresponding to a system characteristic function. The output of a shaped sensor positioned on a finite simply supported beam is given by

$$q(t) = K \sum_{m=1}^{\infty} \int_0^L \Gamma(x) \Psi_m \, dx, \tag{15}$$

where the mode shape is $\Psi_m = \sin(m\pi/L)$. If the shape function $\Gamma(x)$ is chosen equal to a mode shape $\Psi_m(x)$, then, by orthogonality, Eq. (15) predicts that the sensor will only observe that mode. Figure 2 presents an example of a sensor shaped to only observe the $m = 2$ mode of the simply supported beam. Note that the sensor width is largest where the surface strain of the beam is highest. Thus, shaped sensors can provide a time-domain estimate of modal amplitudes of structural response, with minimal signal processing, in contrast to the previous method using arrays of accelerometers. However, once the sensor is shaped, its estimate is fixed, whereas the

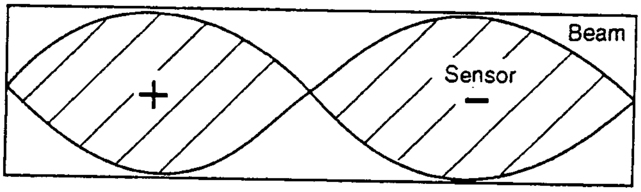


Fig. 2 Shaped distributed piezoelectric strain sensor on a simply supported beam.

output of the accelerometer array can be manipulated by signal processing to observe various modes or waves.

The most commonly used piezoelectric sensing material is PVDF. It can easily be cut or etched into particular shapes. It is readily attached to the surface with double-sided tape or other means. Care must be taken to attach leads to the bottom of the element, and when the shaped sensor has phase changes throughout its segment (as in Fig. 2), each segment output should be wired together with the correct relative polarity in order to give the required sensor output.

4 CONTROL APPROACHES

The function of the controller is to process the error information received from the error sensors in order to calculate control signals so as to cause the controlled or closed-loop system to behave in a required manner. Control approaches can be divided into two main categories: *feedforward* and *feedback* approaches, as discussed by Fuller et al.[3]

In general, feedforward control has found application when the designer has direct access to information about the disturbance signal to the system. On the other hand, feedback control has primarily been applied when the disturbance cannot be directly observed. In this case the control signals are obtained from the sensor(s) whose output is affected by both disturbance source and the control actuator(s). Chapter 84 also has relevant material on the active control paradigms discussed here.

4.1 Feedforward Control

Figure 3 presents a simple, illustrative arrangement of single-channel feedforward control; in this particular case the object is to minimize the vibration of a mechanical system at the error sensor, e. The most common form of electrical feedforward controller H is an adaptive finite impulse response (FIR) filter whose coefficients are updated by a control algorithm in order to drive the error signal to a minimum. A critically important aspect of the feedforward approach is obtaining a reference or training signal that provides a coherent estimate of the disturbance signal and is "fed forward" through the adaptive filter to provide the control signal. The most commonly used update equation for dynamic systems is the time-domain Filtered X version of the Widrow-Hoff least mean squared (LMS) algorithm written in single-channel form as[23]

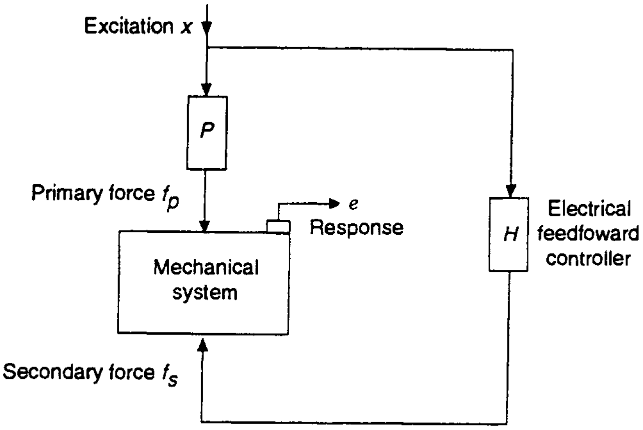


Fig. 3 Simple feedforward control arrangement.

$$w^i(n+1) = w^i(n) - 2\mu e(n) f(n), \tag{16}$$

where $w^i(n)$ is the ith coefficient of the adaptive filter at time n, μ is a convergence parameter, and $e(n)$ is the error signal at time n. The filtered reference signal $f(n)$ is given by

$$f(n) = \sum_{j=1}^{P} c^j x(n-j), \tag{17}$$

where $c^j, i = 1, 2, \ldots, P$, are the weights of a fixed FIR digital filter that models the plant dynamics from the actuator signal to the error signal and $x(n)$ is the reference signal taken directly from the disturbance. Equation (17) essentially accounts for the delay between the control system input and the error sensor output. The adaption equation (16) also assumes no feedback of control signal to the reference signal. The fixed FIR filter is usually constructed before switching on the controller or with a low-level probe signal, as described in Refs. 3 and 4. In essence the Filtered X LMS algorithm is a gradient-descent technique that finds the minimum of a quadratic cost function defined as the squared modulus of the error signal.

The choice of the filter format is dependent upon the noise disturbance. For a narrow-band signal it is necessary to use only two filter coefficients. Control of broadband signals requires a larger filter in terms of coefficients, and causality aspects associated with relative delay through the control path become more important.[3] The Filtered X algorithm can be readily extended to a multichannel configuration, as discussed in more detail in Ref. 24.

Note that the use of a reference signal estimate from the disturbance implies that the system will respond only to information coherent with the reference signal. The system will thus appear open loop to all other disturbances uncorrelated with the error signal. The maximum achievable attenuation of the power of the error signal for a single-input, single-output feedforward control arrangement can be calculated from[3,4]

$$\Delta \mathrm{dB} = -10\log(1-\gamma_{de}^{2}), \qquad (18)$$

where γ_{de}^2 is the coherence between the disturbance and the error signals.

Similar to feedback control, it can be demonstrated that closed loop systems under feedforward control have new eigenvalues and eigenfunctions[25] and that, for narrow-band disturbances, feedforward and feedback approaches are essentially equivalent.[3] Due to convergence time requirements (the adaptive filter has to search for the minimum of the error surface), the feedforward algorithm does not presently work well with transients applied to nonmodelled systems but is suitable for pure tones or steady-state broadband signals. As the control filter is adaptive and "learns" the system by minimizing a cost function, it is not necessary to have an accurate model of the open-loop system behavior from the control actuator to the error sensor. For a single-channel time-domain controller applied at a single frequency, it is only required to estimate the fixed transfer function between the control input and error sensor within an accuracy of ±90° to ensure controller convergence.

The behavior of feedforward controlled systems can also be analyzed using linear quadratic optimal control theory. For the system of Fig. 3 a quadratic cost function is constructed by squaring the modulus of the error signal (or summing the squared modulus of the error signals). The optimal control input is found by setting the gradient of the cost function to zero (i.e., at the minimum).

For a system with L error sensors the quadratic cost function is formed by summing the squared modulus of the error signals,

$$J = \sum_{l=1}^{L} |e_l|^2. \qquad (19)$$

The optimal control inputs to minimize the cost function are found by setting the gradient of the cost function to zero, as described by Nelson and Elliott.[26] For a multiple-input, multiple-output arrangement the optimal vector q_{s0} of control inputs to minimize the cost function is specified by

$$q_{s0} = A^{-1}Z^{H}w_p, \qquad (20)$$

where Z is a matrix of complex transfer impedances relating the control force inputs to the response at the error sensors, the matrix $A = Z^{H}Z$ and w_p is the noise or primary complex field at the error sensors. More details on how to use Eq. (20) as well as calculating the minimum of the cost function can be found in Ref. 26. The above analysis is for a single-frequency, steady-state excitation and represents the maximum achievable performance. For random signals, the actual control performance will depend upon whether the control filters are physically realizable and other related issues, as described in Ref. 4.

4.2 Feedback Control

Figure 4 presents a typical arrangement of single-channel feedback control of motion of a general mechanical system. In this arrangement the control signal is derived from a sensor on the structure and is then "fed back" through a controller (or compensator) to provide a control input. The optimal controller is derived in order to perform a specified control task. Hence, generally speaking, for feedback control the system changes its dynamic characteristics for *all* disturbances applied to it, in contrast to the feedforward approach discussed previously.

The open-loop equation of motion of a single-degree-of-freedom system excited by a disturbance force $f_p(t)$ can be written in the form[3]

$$m\ddot{w}(t) + c\dot{w}(t) + kw(t) = f_p(t), \qquad (21)$$

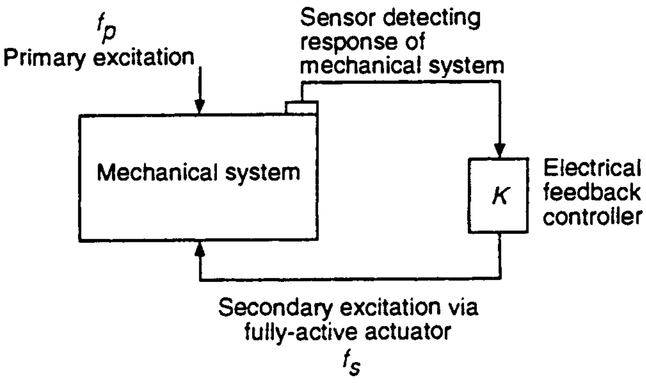

f_p
Primary excitation
Sensor detecting response of mechanical system
Mechanical system
K
Electrical feedback controller
Secondary excitation via fully-active actuator
f_s

Fig. 4 Simple feedback control arrangement.

where m, c, and k are the system mass, damping, and stiffness, respectively. For the mechanical system in Fig. 4 this may correspond to the response of one mode. Extensions to distributed systems are described in Refs. 2 and 3. The open-loop transfer function (i.e., the ratio of the response to the disturbance) of this system can be evaluated using the Laplace transform[3] and is given by

$$G(s) = \frac{W(s)}{F_p(s)} = \frac{1}{ms^2 + cs + k}. \tag{22}$$

We now apply a secondary control force to the structure by feeding back a signal taken from the response of the system through a compensator with a transfer function K to drive the control force $f_s(t)$. The closed-loop transfer function of the system with stiffness control (for example) is now given by

$$\frac{W(s)}{F_p(s)} = \frac{G(s)}{1 + G(s)K(s)} = \frac{1}{ms^2 + cs + (K + k)s}, \tag{23}$$

where s is the Laplace variable and K is the compensator gain or transfer function. Note that the closed-loop system effectively has a new effective stiffness $(K + k)$ and thus new poles.

For transient control the objective is often to increase the damping of the system. In this case the closed-loop poles are moved further into the left s-plane by varying the compensator gain K. However, an important aspect of feedback control is stability, which can be determined by inspection of the position of the closed-loop poles as described in Ref. 27. Too much compensator gain K, for example, can lead to an unstable control system. Often, in practice, the stability is assessed from input–output measurements made on the system before control. The Nyquist stability criterion is then applied to check if the system is stable, as described in Ref. 26. The outcome is that the closed-loop system is stable only if the polar plot of the open-loop frequency response function does not enclose the point $(-1, 0)$ in the complex Laplace plane s. Different forms of the compensator K to the simple example discussed above can be chosen, depending upon the required control performance, and these are discussed in Ref. 3.

If it is possible to achieve a stable feedback system, then the error signal power is approximately reduced by an amount proportional to the loop gain,[4] and the attenuation is given by

$$\Delta\text{dB} = -10 \log \left| \frac{1}{1 + GK} \right|^2. \tag{24}$$

In Eq. (24), $G(s)$ is the transfer function through the plant and $K(s)$ is the transfer function through the compensator.

In general, in order to calculate the optimal controller K, it is required to have an accurate model of the system to be controlled. This can be done through calculations such as finite-element methods or more usually through estimates using system identification approaches as described in Ref. 3. Although multi-input, multi-output feedback control can have a larger implementation requirement than feedforward control, it ensures that the system is controllable for all forms of linear noise disturbances if a realizable controller can be defined.

Extension of feedback control to multi-input, multi-output arrangements can be achieved using state space methods as outlined in Refs. 27 and 3. Other important factors affecting the performance and stability of feedback controllers are reviewed in Ref. 4.

5 EXAMPLE APPLICATIONS OF ACTIVE VIBRATION CONTROL

The following examples illustrate some of the wide diversity of present applications of active vibration control. Other implementations of active vibration control are discussed in Ref. 3.

5.1 Active Isolation of Vibration

A classic approach to reducing unwanted machinery noise and vibration is to control these disturbances at transmission bottle necks (e.g., the machinery mount). Active vibration isolation has been studied by a number of authors, as summarized in Ref. 3, and a few representative examples will be presented here.

There are four possible configurations for active isolation: The two most common are *series*, in which the actuator opposes the load path, and *parallel*, where the active force is applied in parallel with the load path. As discussed by Scribner et al.,[28] if the actuator is placed in parallel, it must overcome the stiffness of the passive mount and the actuator deflection requirements are relatively larger. If the actuator is in series, then while the deflection requirements are lower than the parallel implementation, the actuator must carry the load and, as will be seen later, the force requirements at resonance are higher. Two other, less common implementations are when the active force is directly applied to the exciting system and when the active or secondary force is applied directly to the receiving structure at the mount location. This latter implementation is termed *opposed*. Both implementations have the disadvantage of requir-

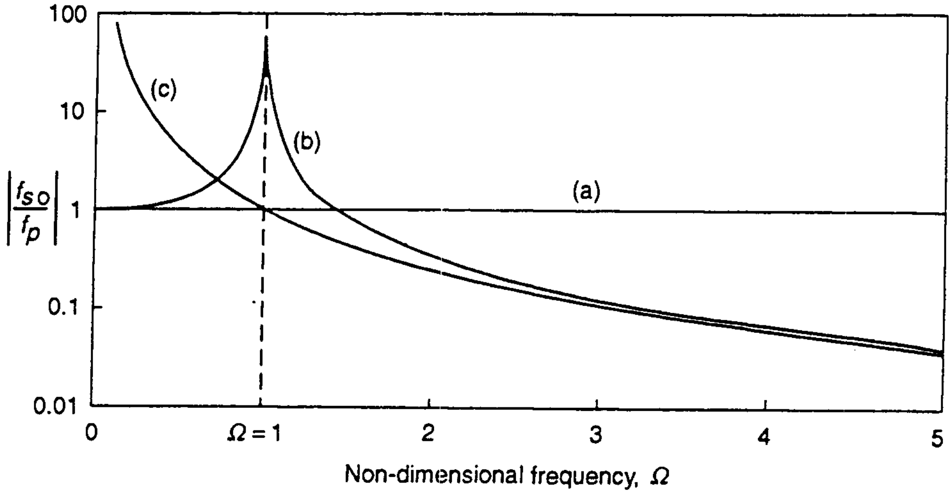


Fig. 5 Control force required for (*a*) direct, (*b*) opposed, and (*c*) parallel active vibration isolation systems.[29]

ing some form of inertial exciter (e.g., an electrodynamic shaker that is provided with a mass against which it can react).

Figure 5 presents the magnitude of the ratio of the secondary control force f_{s0} to the primary disturbance force f_p required for a "parallel" and "opposed" implementation calculated by Nelson et al.[29] for a lumped-parameter single-degree-of-freedom mass–spring system. The nondimensional frequency $\Omega = \omega/\omega_n$, where ω_n is the natural frequency of the mass–spring system on a rigid base. Also shown is the relatively trivial case of when the control force is directly applied to the exciting mass and thus exactly equals the disturbance magnitude in order to drive the motion to zero. The results of the opposed case are similar in trend to the series implementation, which is discussed in much more detail in Ref. 30. The results show that at resonance the force requirements of the opposed system are much larger while below resonance the converse is time. Well above resonance the force requirements of both arrangements are similar. For control systems designed to operate near the plant resonance the above characteristics usually imply that electrodynamic actuators (i.e., relatively large deflection, lower force) are used for the parallel implementation while piezoelectric ceramic actuators (i.e., relatively larger force, low deflection) are used for the series arrangement.

Scribner et al. has theoretically and experimentally studied the use of piezoelectric ceramic actuators in a "series" implementation of active isolation using a classical feedback approach[28] focused upon sinusoidal disturbances. Figure 6 presents typical values of open-loop and closed-loop behavior of the ratio of transmitted to applied force versus frequency. In this example the compensator natural frequency was set to 877 Hz and a large enhancement in transmitted force is observed near this design frequency. However, in the frequency range of $825 \leq f \leq 850$ Hz, high attenuation in the transmitted force is achieved. Scribner et al.[28] also demonstrated how the controller could be implemented in a self-tuning implementation so as to track a disturbance frequency, even through structural resonance.

Nelson et al.[3,29] have also analytically and experimentally investigated narrow-band isolation of the vibration of a rigid raft from a flexible receiving structure using multiple active isolators configured in the parallel arrangement. In this case the active inputs were provided by four electromagnetic coil actuators used in conjunction with a multichannel adaptive LMS feedforward control approach. The error signals were taken from either four accelerometers located adjacent to the mount points or eight accelerometers distributed over the surfce of the receiving structure. The results demonstrate that the use of eight accelerometers was more effective than four in attenuating the total energy of the receiving structure over much of the frequency range. This behavior illustrates the advantage in feedforward control of using more sensors than actuators; a "square" feedforward control system with equal number of actuators and sensors will always drive the response to zero at the error sensors. However, it is often more beneficial to use more error sensors than actuators when global reduction is required. The results presented in Ref. 3 also show that the passive isolation (i.e., no active control) produced an *increase* in the receiving structure vibration at some of its resonance frequencies. This characteristic illustrates the importance of the receiving structure dynamics on the isolation performance.

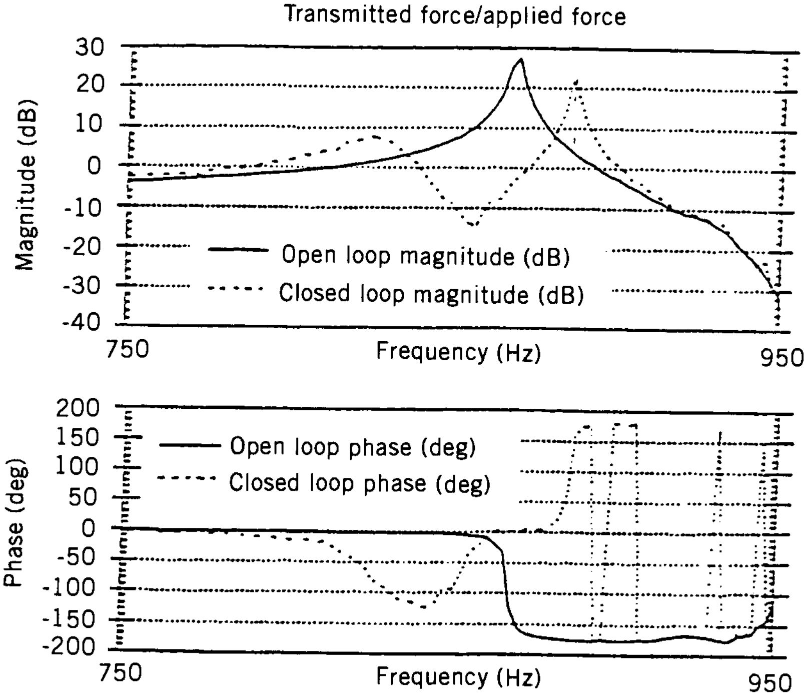


Fig. 6 Performance of a series active vibration isolation system.[28]

5.2 Active Control of Power Flow in Beams

Many structures are composed of basic elements that can be usefully approximated as beams and can transmit unwanted vibrations. The vibrational energy eventually spreads out through the ever increasing complexity of the structure making it more difficult to control. Thus direct reduction of vibrational power flow in the support beams is an area of great interest. In effect this problem is a subset of the active isolation problem discussed above.

Figure 7 shows an experimental arrangement where a semi-infinite beam is driven by a shaker at a single frequency positioned at one end (the disturbance). Piezoelectric ceramic actuators and sensors are positioned on the beam as shown and are wired so as to produce and sense pure bending, as discussed in Section 2.2. The out-of-plane velocity is measured with a laser vibrometer. The control objective is to minimize the power flow transmitted in the beam. Also shown in Fig. 7 is a schematic of a single-channel feedforward control implementation designed to minimize the square of the output of the piezoceramic error sensor which, for a semi-infinite thin beam system, is proportional to total power flow.[31]

The control algorithm used in the arrangement of Fig. 7 was the single-channel Filtered X paradigm of Eq. (16). The controller, indicated by the dashed boundary, was implemented on a TMS320C25 digital signal processing chip with a reference signal taken directly from the disturbance oscillator. Net power flow in the beam was also calculated using the finite-difference equation (12) with the laser vibrometer data. Figure 8 presents a typical result for the measured power flow before and after control is implemented.[31]

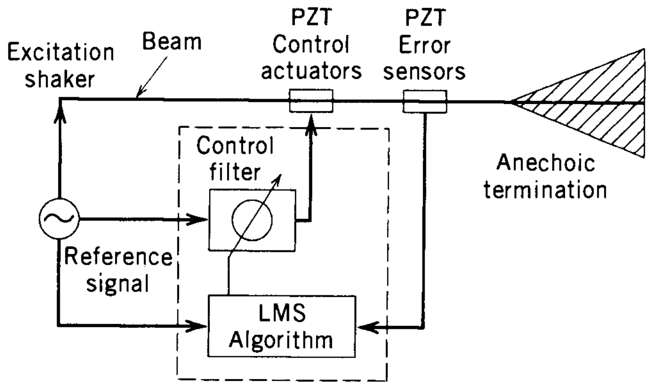


Fig. 7 Experimental arrangement and control block diagram for feedforward active reduction of narrow-band power flow in a beam with piezoelectric actuators and sensors.[31]

Figure 8 shows that when control is applied, the power flow is attenuated by approximately 35 dB at the error sensor. The results of Fig. 8 are net power flow and thus

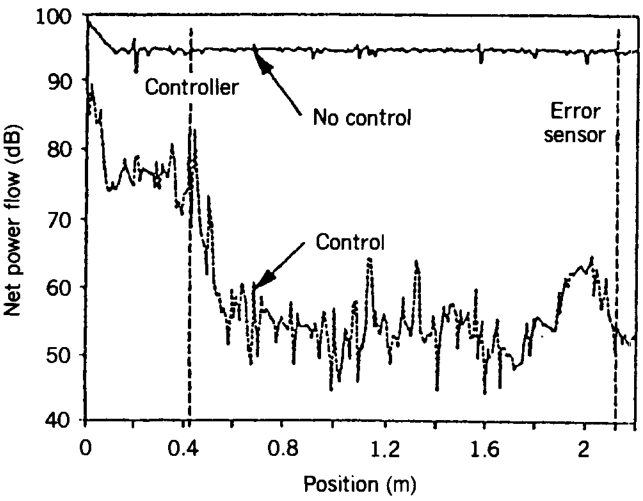


Fig. 8 Power flow distribution with and without control for the system of Fig. 7; f = 302 H.[31]

imply that power flow is highly attenuated at all points downstream of the error sensor. Two other observations of interest are found in the figure. First, the input power at the drive point is attenuated by around 20 db. Thus the control is achieved mainly by the control actuator creating a reactive discontinuity at its location, reflecting waves back to the drive point and lowering its resistive input impedance. Second, the net power falls past the distributed piezoelectric control actuator, and this implies that the control actuator has also *absorbed* some vibrational power. Extension of the above approach to control of power flow in finite beams, simultaneous control of flexural and extensional waves in beams, and broadband disturbances are discussed in Refs. 31, 32, and 33, respectively. The method has also been successfully applied to attenuation of power flow in struts used to mount the engine and gearbox of a helicopter to the fuselage structure.[34]

5.3 Distributed Active Vibration Control of Plates

The control of vibration of distributed structural elements such as plates is also an important problem. Figure 9 is an experimental arrangement used by Rubenstein et al.[35] to implement linear quadratic Gaussian state space feedback control[3] of the vibration of a simply supported plate. The response of the system is sensed using an array of 12 accelerometers of the plate. The output of the accelerometers is passed through modal filters to obtain modal amplitudes and then through a Kalman filter to obtain the states of the system. The optimal controller $K(\hat{Q})$ can then be obtained using a Riccati equation solution of the state space feedback control arrangement.[3] In this experiment the steel panel was 0.6 × 0.5 m and 3 mm thick. The disturbance and control were implemented by electrodynamic shakers. Figure 10*a* shows a typical closed-loop response for mode 1 when the disturbance is narrowband, continuous excitation, while Fig. 10*b* gives closed-loop response to a transient disturbance applied at near $t = 0.65$ s. In both cases high modal attenuation of the disturbance is achieved. Thus state space feedback methods show much potential for distributed active vibration control of a wide range of input disturbances. Note that state space feedback control implies complete control of specified states of the system as opposed to feedforward control, which usually controls a direct output of the plant at the error sensors (displacement, velocity, etc.). As discussed previously, the state space approach also guarantees control of the system to a wide range of disturbances without requiring access to a reference signal.

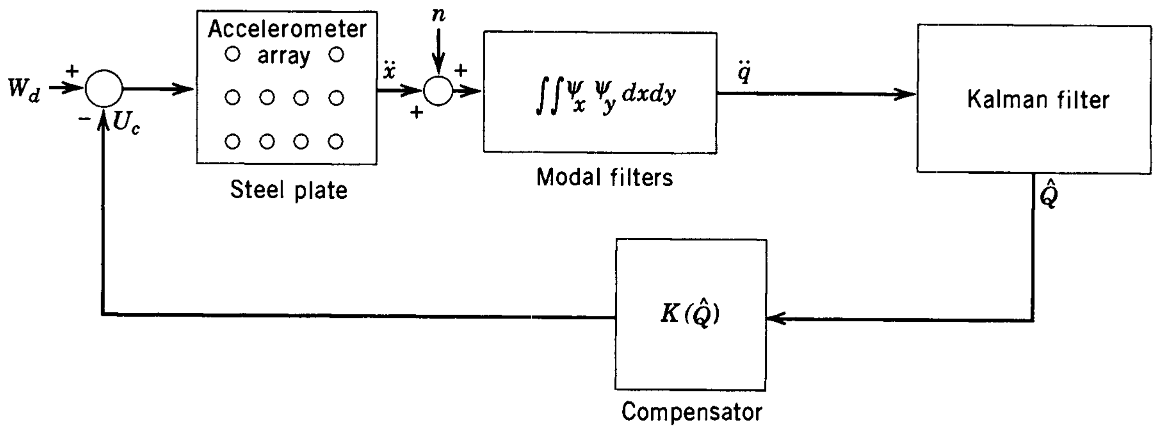


Fig. 9 Linear Quadratic Gaussian (LQG) feedback control system for active reduction of plate vibrations.[35]

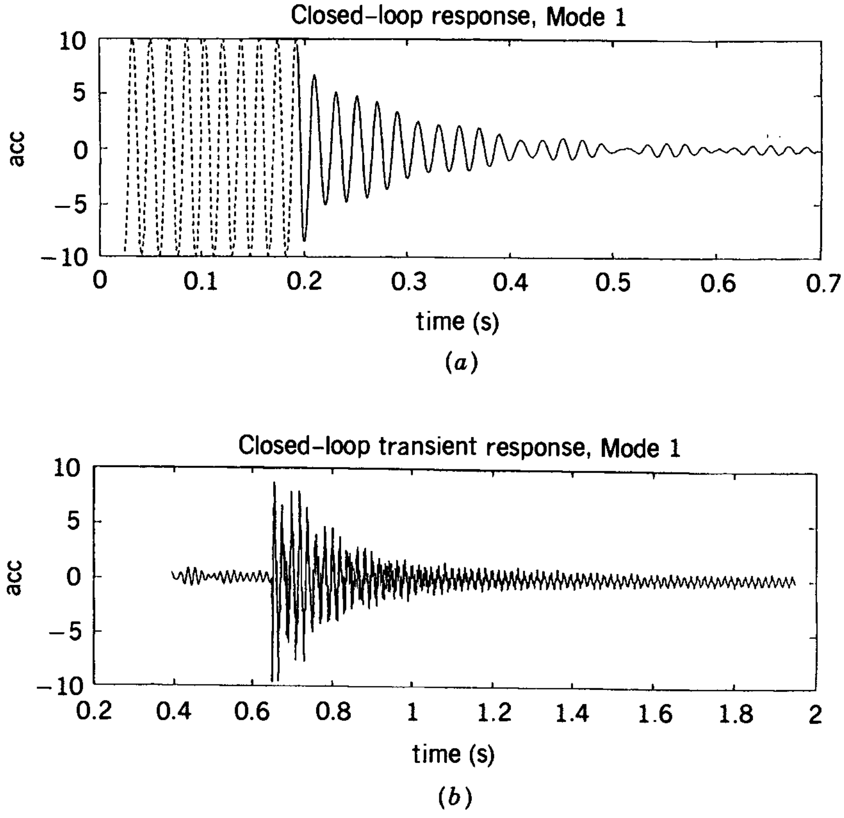


Fig. 10 Closed-loop response of the system of Fig. 10 for mode 1 with (*a*) narrow-band steady disturbance and (*b*) transient disturbance at $t = 0.65$ s.[35]

5.4 Active Structural Acoustic Control

Vibrating structures often radiate or transmit unwanted sound. An effective technique to control the structural sound radiation is to apply control inputs directly to the structure while minimizing the radiated sound or components of structural motion associated with the sound radiation.[3, 36] The technique called active structural acoustic control (ASAC) thus falls in the class of general active vibration control techniques. Figure 11 shows an experimental arrangement in which the sound radiation from a simply supported plate is to be minimized.[22] The noise disturbance is provided by a point force shaker attached to the plate and driven by a pure tone. The control inputs are multiple colocated piezoelectric ceramic actuators bonded to the plate surface and wired to produce pure bending. The error sensors are either microphones in the radiated far field or shaped PVDF sensors attached to one side of the plate. The shaped PVDF sensors are rectangular strips positioned on the plate, as discussed in Section 3.3, so as to observe only the odd–odd modes (the efficient modal radiators at low frequencies). The control arrangement is the multichannel feedforward implementation of the Filtered X algorithm discussed previously. The reference signal is taken from the noise input signal generator and is filtered as outlined in Eq. (17) before it is used in the LMS update equation, as described in Section 4.1.

Figure 12 presents results at 349 Hz, which is near the (3, 1) resonant frequency of the plate. The dotted line is for a case where three piezoelectric ceramic actuators are used in conjunction with three error microphones in the radiated far field, and it is apparent that global attenuations of approximately 25 dB are achieved. The dashed line is a similar case except that now two piezoceramic actuators are used in conjunction with two PVDF strips arranged in the x- and y-direction on the plate, as outlined in Section 3.3. The results again demonstrate good global control of approximately 20 dB in the radiated sound levels. These results and others discussed in Ref. 3 demonstrate the potential of using ASAC to reduce radiated sound. For the ASAC technique there are two main mechanisms of control.[16] *Modal suppression* corresponds to a direct reduction of the dominantly radiating modes or a fall in the plate wavenumber spectrum. *Modal restructuring* corresponds to the residual plate response having a lower overall radiation efficiency while its total magnitude of response is reduced

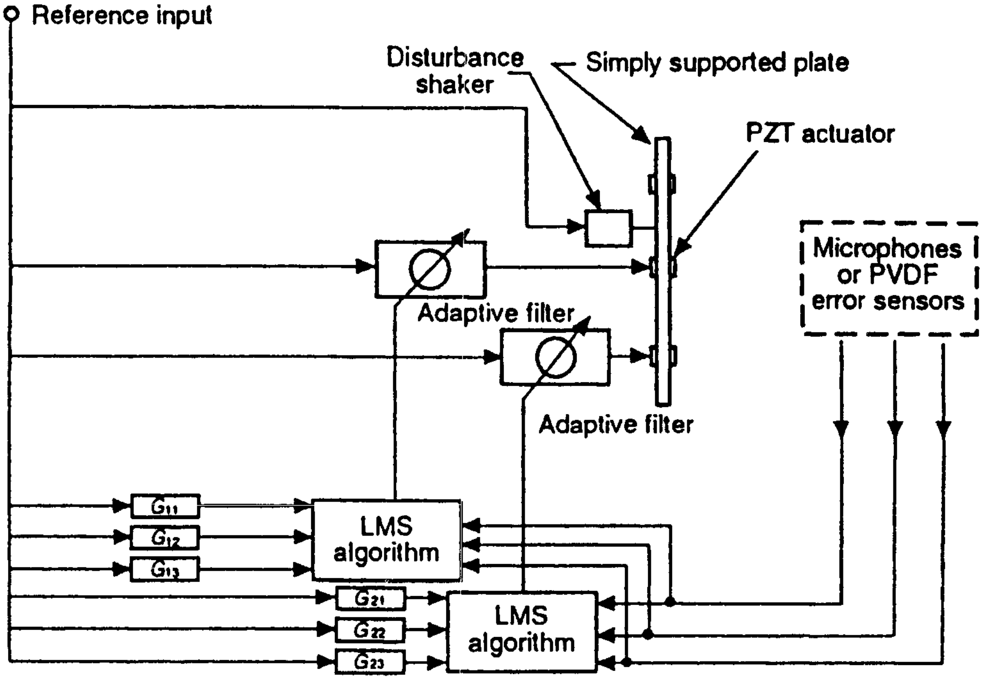


Fig. 11 Experimental arrangement and control block diagram for active reduction of sound radiation from a plate (ASAC).[22]

only slightly or may even increase. This second case corresponds to attenuation of only supersonic components of the plate wavenumber spectrum while subsonic, nonradiating components are left unaffected. Thus the ASAC technique takes maximum advantage of the natural structural–acoustic coupling behavior to reduce the dimensionality of the controller. The technique has been successfully applied to the control of interior noise in aircraft and the reduction of transformer radiated noise, as discussed in Refs. 3 and 4.

From a system point of view, when multiple integrated actuators and sensors such as piezoelectric devices are used in conjunction with a non-model-based, adaptive controller, the application (such as ASAC discussed above) can be seen to be closely related to the field of *smart*, *adaptive*, or *intelligent* structures (see Ref. 37 for a review of this field). Developments in the field of smart structures are likely to be used in advanced active vibration control and ASAC because of the inherent compactness, light weight, and adaptability of this class of systems.

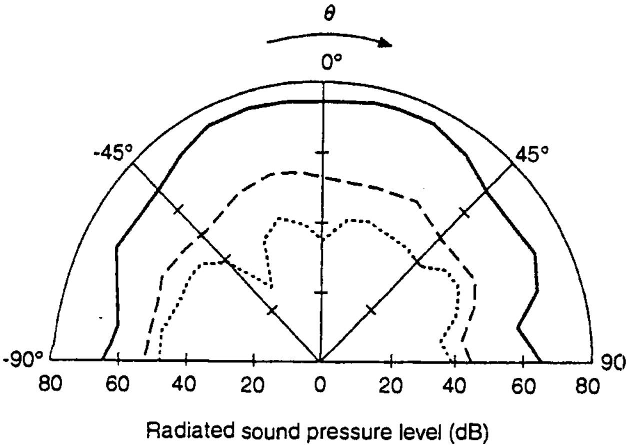


Fig. 12 Active structural acoustic control performance for the system of Fig. 11: (——) uncontrolled (-------) controlled with PVDF structural error sensors; (.........) controlled with far-field error microphones; $f = 349$ Hz.[22]

REFERENCES

1. L. L. Beranek, *Noise and Vibration Control*, INCE, Washington, DC, 1988.
2. L. Meirovitch, *Dynamics and Control of Structrues*, Wiley, New York, 1990.
3. C. R. Fuller, S. J. Elliott, and P. A. Nelson, *Active Control of Vibration*, Academic, London, 1996.
4. C. R. Fuller and A. H. Von Flotow, "Active Control of Sound and Vibration," *IEEE Control Systems Mag.*, December, 1995, pp. 9–19.
5. Ling Dynamic Systems, *User Manual*, Ling Dyamic Systems, Herts, England, 1973.

6. L. Cremer, M. Heckl, and E. E. Ungar, *Structural-Borne Sound*, Springer-Verlag, Berlin, 1988.
7. A. J. Moulson and J. M. Herbert, *Electroceramics: Materials, Properties, Applications*, Chapman and Hall, London, 1990.
8. E. F. Crawley and J. de Luis, "Use of Piezoelectric Actuators as Elements of Intelligent Structures," *AIAA J.*, Vol. 25, No. 10, 1989, pp. 1373–1385.
9. R. L. Clark, C. R. Fuller, and A. L. Wicks, "Characterization of Multiple Piezoceramic Actuators for Structural Excitation," *J. Acoust. Soc. Am.*, Vol. 90, No. 1, 1991, pp. 346–357.
10. E. K. Dimitriadis, C. R. Fuller, and C. A. Rogers, "Piezoelectric Actuators for Distributed Vibration Excitation of Thin Plates," *J. Vib. Acoust.*, Vol. 113, 1991, pp. 100–107.
11. C. M. Jackson, H. J. Wagner, and R. J. Wasilewski, "55-Nitinol-The Alloy with a Memory: Its Physical Metallurgy, Properties and Applications," NASA-SP-5110, Washington, DC, 1972.
12. C. A. Rogers, "Active Vibration and Structural Acoustic Control of Shape Memory Alloy Hybrid Composites: Experimental Results," *Proceedings of International Congress on Recent Developments in Air-And Structure-Borne Sound and Vibration*, Auburn University, AL, 1990, pp. 695–707.
13. C. G. Miller, "High Force, High Strain, Wide Bandwidth Linear Actuators Using the Magnetostrictive Material Terfenol-D," in C. A. Rogers and C. R. Fuller (Eds.), *Proceedings of Recent Advances in Active Control of Sound and Vibration*, (supplement) Technomic, Lancaster, PA, 1991.
14. J. E. Stangroom, "Electrorheological Fluids," *J. Phys. Tech.*, Vol. 14, 1983, pp. 290–296.
15. M. V. Gandhi and B. S. Thompson, "Dynamically-Tunable Smart Composite Featuring Electro-rheological Fluids," *Proceedings of the SPIE Conference on Fiber Optic Smart Structures and Skins II*, Boston, MA, 1989, pp. 294–304.
16. C. R. Fuller, C. H. Hansen, and S. D. Snyder, "Active Control of Sound Radiation from a Vibrating Rectangular Panel by Sound Sources and Vibration Inputs: An Experimental Comparison," *J. Sound Vib.*, Vol. 145, No. 2, 1991, pp. 195–215.
17. G. P. Gibbs, C. R. Fuller, and R. J. Silcox, "Active Control of Flexural and Extensional Power Flow in Beams using Real Time Wave Vector Sensors," *Proceedings of the 2nd Conference on Recent Advances in Active Control of Sound and Vibration*, Technomic, Lancaster, PA, 1993, pp. 909–925.
18. K. J. Astrom and B. Wittenmark, *Computer Controlled Systems*, Prentice-Hall, Englewood Cliffs, NJ, 1984.
19. J. M. Downing and K. P. Shepherd, "Power Flow in Beams Using a 5-Accelerometer Probe," *Proceedings of Noise-Con 88*, Purdue University, IN, 1988, pp. 335–340.
20. J. Scheuren, "Active Control of Bending Waves in Beams," *Proceedings of Inter-Noise 85*, Munich, 1985, pp. 591–595.
21. C. K. Lee, W. W. Chiang, and T. C. O'Sullivan, "Piezoelectric Modal Sensors and Actuators Achieving Critical Damping on a Cantilever Plate," *AIAA Paper*, No. 89-1390, 1989.
22. R. L. Clark and C. R. Fuller, "Modal Sensing of Efficient Radiators with PVDF Distributed Sensors in Active Structural Acoustic Approaches," *J. Acoust. Soc. Am.*, Vol. 91, No. 6, 1992, pp. 3321–3329.
23. B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1985.
24. S. J. Elliott, I. M. Stothers, and P. A. Nelson, "A Multiple Error LMS Algorithm and Its Application to the Active Control of Sound and Vibration," *IEEE Trans. Acoust., Speech and Signal Process.*, ASSP-35, 1987, pp. 1423–1434.
25. R. A. Burdisso and C. R. Fuller, "Theory of Feedforward Controlled Systems Eigenproperties," *J. Sound Vib.*, Vol. 153, No. 3, 1992, pp. 437–451.
26. P. A. Nelson and S. J. Elliott, *Active Control of Sound*, Academic, London, 1992.
27. W. L. Brogan, *Modern Control Theory*, Prentice-Hall, Englewood Cliffs, NJ, 1982.
28. K. B. Scribner, L. A. Sievers, and A. H. von Flotow, "Active Narrow Band Vibration Isolation of Machinery Noise From Resonant Substructures," *J. Sound Vib.*, Vol. 167, No. 1, 1993, pp. 17–40.
29. P. A. Nelson, M. D. Jenkins, and S. J. Elliott, "Active Isolation of Periodic Vibrations," *Proceedings of Noise-Con 87*, Penn State University, State College, PA, 1987, pp. 425–430.
30. A. H. von Flotow, "An Expository Overview of Active Control of Machinery Mounts," *Proceedings of the 27th Conference on Decision and Control*, Austin, TX, 1988.
31. G. P. Gibbs and C. R. Fuller, "Experiments on Active Control of Vibrational Power Flow Using Piezoceramic Actuators and Sensors," *AIAA J.*, Vol. 30, No. 2, 1992, pp. 457–463.
32. C. R. Fuller, G. P. Gibbs, and R. J. Silcox, "Simultaneous Active Control of Flexural and Extensional Waves in Beams," *J. Intell. Mat. Systems Struct.*, Vol. 1, No. 2, 1990, pp. 235–247.
33. S. J. Elliott and L. Billet, "Adaptive Control of Flexural Waves Propagating in a Beam," *J. Sound Vib.*, Vol. 163, 1993, pp. 295–310.
34. S. J. Elliott, T. J. Sutton, M. J. Brennan, and R. J. Pinnington, "Vibration Reduction by Active Wave Control in a Strut," *Proceedings of the IUTAM Symposium on the Active Control of Vibration*, Bath, England, 1994, pp. 1–8.
35. S. Rubenstein, W. R. Saunders, G. K. Ellis, H. H. Robertshaw, and W. T. Baumann, "Demonstration of a LQG Vibration Controller for a Simply Supported Plate," in C. A. Rogers and C. R. Fuller (Eds.), *Proceedings of Recent Advances in Active Control of Sound and Vibration*, Technomic, Lancaster, PA, 1991, pp. 618–630.
36. C. R. Fuller, "Analysis of Active Control of Sound Radiation from Elastic Plates by Force Inputs," *Proceedings of Inter-Noise 88*, Avignon, France, 1988, pp. 1061–1064.
37. B. K. Wada, J. L. Fanson, and E. F. Crawley, "Adaptive Structures," *Mech. Eng.*, November, 1990, pp. 41–46.

PART VII

STATISTICAL METHODS IN ACOUSTICS

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker

Copyright © 1997 John Wiley & Sons, Inc.

76

INTRODUCTION

RICHARD H. LYON

1 INTRODUCTION

The following two chapters deal with some particular uses of statistics in acoustics. They have to do with the response of systems—sound fields and structures—to excitation. The main point to be made is that we can consider the excitation and/or response of systems to be random (or both).

Studies of sound fields in rooms have for a long time been considered statistical descriptions of the sound field, even when the excitation and response are deterministic. Modal descriptions used concepts such as modal density and resonance frequency spacing statistics. Wave descriptions, particularly, called ray acoustics at "high frequencies" have been statistical and have used descriptors such as mean free path or diffuseness. If the excitation is random, so much the better for the statistics.

The development of statistical ideas applied to structural response has proceeded in quite a different way. Vibration problems are often dominated by the very lowest modes of a structure. Consequently, engineers have thought of structures as deterministic, but the excitation and response of structures is often random.

2 RANDOM RESPONSE OF STRUCTURES AND SOUND FIELDS

The details of the random response of structures have been more important in structures than for sound fields. The room acoustician has usually been happy to know the frequency spectrum of the sound pressure and possibly its variance as the source and/or receiver move from one location to another. The structural engineer wants to know how likely a certain exceedance will occur (will one part of the structure bang into another?) Or, he or she may be calculating fatigue life and be interested in certain high-order moments of the response. We could describe the acoustician as interested in simple statistics on complicated systems, while the structural engineer is interested in complicated statistics on simple systems.

This dichotomy more or less describes the situation in the late 1950s and early 1960s. But that changed with increasing concern about the vibrational characteristics of large and complex structures, large launch rockets, and the broadband noise of engines that could not be analyzed by the few fundamental mode procedures. The sound on ships produced by their machinery, or by the turbulent flow over null structures could not be understood by conventional models.

It became apparent that new analytical techniques were needed to deal with large complex structures excited by broadband noise. In the early 1960s, new techniques using finite-element methods (FEMs) appeared to offer a solution, and large-scale computer programs such as NASTRAN were developed. As computers became more powerful over time, the FEM codes and the problems tackled by them increased in complexity. But it is safe to say that the goal of solving realistic large-scale structural response problems with these methods continues to be a tantalizing but unachieved goal.

3 STATISTICAL ENERGY ANALYSIS

At about the same time, a different approach was developed in which both sound fields and structures were statistically described. Although room acoustics had long

---

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker
Copyright © 1997 John Wiley & Sons, Inc.

used such descriptions, such an idea was not easily assimilated by structural engineers, who, after all, have detailed mechanical drawings for their structures. How could something built to such a prescription be considered statistical? This new approach was termed statistical energy analysis (SEA) and has since become a fairly standard, if somewhat misunderstood, tool in the analyst's bag of tricks.

But there is a fundamental conceptual issue in population statistics applied to system response. Statistical energy analysis and other related disciplines such as statistical room acoustics and statistical mechanics employ the ergodic hypothesis. Simply stated, this means that any one member of the population will obey the statistics of the population as a whole as realizations of its response are repeated. The practical consequence is that one can calculate population statistics using the SEA model and apply the results to a particular individual member of the population, to wit, the one sitting in the laboratory. This assumption, while reasonable, cannot be fully justified for constructed systems like structures and rooms.

4 NEW DEVELOPMENTS

Just as structural systems have had to employ a viewpoint traditionally associated with room acoustics, new developments in room acoustics have led to analysis techniques developed for structural dynamics. Signal processing applications in machine diagnostics has led to an interest in the phase properties of structural transfer function and to the group delay in particular. New room acoustics interests in environment simulation for sound reproduction and in reverberation suppression in group telephony require a much more detailed understanding of the statistical properties of sound fields. Coherence and group delay statistics of small- and intermediate-size rooms are matters of importance.

These new concerns for the modeling and analysis of the phase properties of transfer functions has led to a new methodology we might call statistical phase analysis (SPA). Some readers will be struck by the notion that SEA and SPA should be related by the fact that energy and phase are conjugate variables in classical mechanics (like displacement and momentum). Certainly, the analysis of group delay statistics has been shown to have some very close parallels to response amplitude calculations.

It will be clear to the reader that the methods and models to be applied to sound fields and structures have become closer. It would be a mistake, however, to conclude that the results from one area can be assumed to be directly applicable to the other, since there are some important differences in system parameters. The fact that structures tend to be one and two dimensional (beam- or platelike) while rooms are three dimensional is one difference. Another is the fact that sound waves in rooms are nondispersive while the bending waves are dispersive. One consequence of these differences is that a very important statistical determinant, the system modal overlap, in structures tends to be small while it tends to be large in rooms.

The discussions in the following two chapters will add flesh to the bones of this introductory discussion.

77

RESPONSE STATISTICS OF ROOMS

MIKIO TOHYAMA

1 INTRODUCTION

This chapter describes the statistical characteristics of sound fields in rooms. The reverberation behavior is described by introducing the concept of mean free path[1] into the sound reflection process, similar to the random motion theory of gas. The sound path from a source to a listener is, in principle, predictable in a deterministic way in a closed space. The time history of the sound reflections is expressed by the impulse response between a source and a receiving point. The impulse response, however, is very sensitive to the locations of both the source and receiver even in a simply formed rectangular room. The variabilities of the impulse responses are difficult to predict in any practical sense. Thus, in many cases, it may be more appropriate to assume that the samples of those impulse responses due to different locations make an ensemble of stochastic processes.

The reverberation decay properties in temporal statistics, the transfer functions in the frequency-domain statistics, and spatial distribution statistics are described. Modern reverberation formulas are expressed as well as the conventional ones. The transfer function statistics are extended into the phase analysis based on the distribution of poles and zeros in the frequency domain. The transfer function phase is important in sound field control such as active noise control. Spatial correlations of reverberation sounds are described as the spatial sampling statistics. These statistics are fundamental for acoustic measurement and design in a reverberant space, sound field control, and machine diagnostics using transfer functions.

2 REVERBERATION SOUND IN ROOMS

Reverberation is a transient and nonstationary response of sound in rooms. The reverberation process is fairly slow; thus we can describe the process as a random process.

2.1 Multinomial Distribution of Reflections and Eyring's Formula

The reverberation behavior is described by introducing the concept of mean free path[1] into the sound reflection process, similar to the random motion theory of gases. Suppose that the boundaries of a room are composed of K different surfaces, α_k is the absorption coefficient of the kth surface, and n_k is the number of reflections that occur when a sound ray travels from the kth surface in the time interval t. The reverberation process is written as

$$E_{(n_1,n_2,\dots,n_K)}(t) = E_0 \prod_{k=1}^{K} (1-\alpha_k)^{n_k} \qquad (\mathrm{J/m^3}), \qquad (1)$$

where E_0 is the energy density in the room at steady state ($t = 0$, i.e., the time when the sound source stops). The ensemble-averaged reverberation process is expressed by

$$E(t) = E_0 \sum_{n_1,n_2,\dots} p(n_1,\, n_2,\dots,n_k) \prod_{k=1}^{K} (1-\alpha_k)^{n_k} \qquad (\mathrm{J/m^3}), \qquad (2)$$

where $p(n_1, n_2, \dots, n_k)$ is the joint probability distribution for the n_k. Following Kuttruff,[2] if we assume a multinomial distribution, we obtain Eyring's decay for-

---

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker

Copyright © 1997 John Wiley & Sons, Inc.

mula:

$$E(t) = E_0 \exp\left[\frac{cS}{4V}\, t \ln(1 - \overline{\alpha})\right], \qquad (\mathrm{J/m^3}), \qquad (3)$$

where $\overline{\alpha}$ is the averaged absorption coefficient and the number of reflections N is given by

$$N = \frac{cSt}{4V}, \qquad (4)$$

where $4V/S$ denotes the mean free path for the diffuse-field condition, V is the room volume (in cubic metres), S is the surface area (in square metres) of the room, and c is the sound speed (in metres per second). Thus the reverberation time is given by

$$T_R \approx 0.161 \frac{V}{-\ln(1 - \overline{\alpha})S} \qquad (\mathrm{s}). \qquad (5)$$

2.2 Poisson Distribution and Sabine's Formula

If we assume independent Poisson distributions for the n_k with the average $\overline{n}_k$,[3]

$$\overline{n}_k = \frac{cS_k}{4V}\, t, \qquad p(n_1, n_2, \ldots, n_k) = \prod_{k=1}^{K} \left(\frac{\overline{n}_k^{\,n_k}}{n_k!}\right) e^{-\overline{n}_k}, \qquad (6)$$

then the historically important Sabine formula is obtained from Eq. (2) as

$$E(t) = E_0 \exp\left(-\frac{c}{4V}\, t \sum_k \alpha_k S_k\right) \qquad (\mathrm{J/m^3}). \qquad (7)$$

We can also obtain the reverberation time T_R,

$$T_R = \frac{4V}{c\overline{\alpha}S} (\ln 10^6) \approx 0.161 \frac{V}{\overline{\alpha}S} \qquad (\mathrm{s}). \qquad (8)$$

2.3 Decay Formula for a Two-Dimensional Field

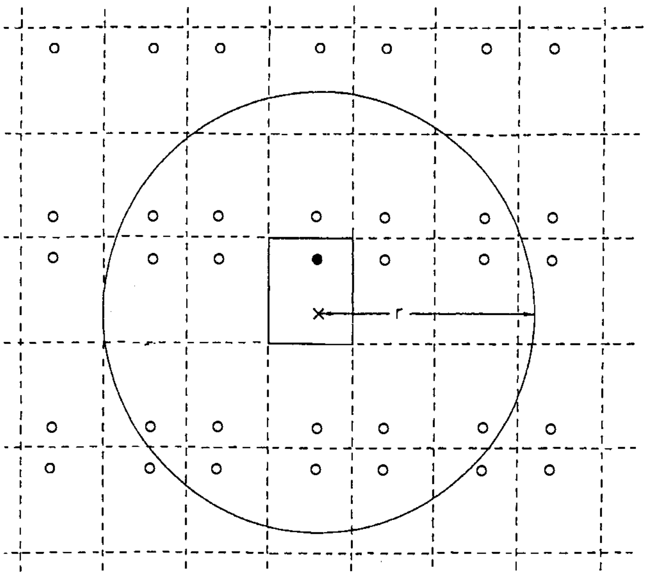


Fig. 1 Two-dimensional array of image sources. (From M. Tohyama, *The Nature and Technology of Acoustic Space*, Academic, London, 1995.)

When the image sources are arranged in a two-dimensional plane including a sound source (Fig. 1), the reverberant energy decay in the two-dimensional field is expressed as a hyperbolic type of function.[4] The average number of the image sources from which the "reflected" impulse waves travel to a receiving point in a time interval dt after t seconds from the arrival of the direct sound is given by

$$n_{2\mathrm{d}}(t) = \frac{2\pi c^2 t}{S_{2\mathrm{d}}}\, dt, \qquad (9)$$

where ct is the distance (in metres) between the receiving point and the location of the image source and $S_{2\mathrm{d}}$ denotes the area (in square metres) of the two-dimensional space. The average number of reflections $N_{2\mathrm{d}}$ is given by

$$N_{2\mathrm{d}} = \frac{ct}{m_{2\mathrm{d}}} = \frac{L_{2\mathrm{d}} ct}{\pi S_{2\mathrm{d}}}, \qquad m_{2\mathrm{d}} = \frac{\pi S_{2\mathrm{d}}}{L_{2\mathrm{d}}} \qquad (\mathrm{m}) \quad (10)$$

at t (seconds) after the direct sound arrives at the receiving point, where $m_{2\mathrm{d}}$ denotes the mean free path in a two-dimensional diffuse field[1] and $L_{2\mathrm{d}}$ (metres) shows the surrounding length of the two-dimensional space. The reverberant decay curve is expressed as

$$E_{2\mathrm{d}}(t) = \int_t^{\infty} \frac{P}{2S_{2\mathrm{d}}} \frac{dt}{t} \exp\left[\ln(1 - \alpha) \frac{cL_{2\mathrm{d}} t}{\pi S_{2\mathrm{d}}}\right] \qquad (\mathrm{W/m^2}), \qquad (11)$$

where P denotes the power output of the point source. Equation (11) is asymptotically written as

$$E_{2d}(t) \to \begin{cases} \displaystyle\int_t^{\infty} \frac{dt}{t}, & (t \to 0), \\ \displaystyle\int_t^{\infty} \exp\left[\ln(1-\alpha)\,\frac{cL_{2d}t}{\pi S_{2d}}\right] dt & (t \to \text{large}). \end{cases} \tag{12}$$

The reverberation time determined by the *later part of the energy decay* becomes

$$T_{R,2d} = \frac{0.128S_{2d}}{-\ln(1-\alpha)L_{2d}} \quad \text{(s)}. \tag{13}$$

2.4 Decay Characteristics in a Nearly Two-Dimensional Field

Suppose that the sound field is assumed to be composed of both tangential waves and oblique waves that are close to tangential waves, called here nearly tangential waves. The reverberation process has a typical frequency characteristic in the midfrequency range, since the three-dimensional field is dominant at low frequencies while the two-dimensional field is significant at high frequencies.[5,6]

Average Number of Reflections and Wavenumber Constant The average number of reflections of an oblique wave in a rectangular room, ν, is given by[7]

$$\begin{aligned} \nu &= \frac{c(L_yL_z|\alpha| + L_zL_x|\beta| + L_xL_y|\gamma|)}{V} \\ &= \nu_x + \nu_y + \nu_z \quad (\text{s}^{-1}), \end{aligned} \tag{14}$$

where L_x, L_y, L_z are the lengths of the sides of the rectangular room (in metres); α, β, γ are direction cosines,

$$\alpha = \frac{k_x}{k_0}, \qquad \beta = \frac{k_y}{k_0}, \qquad \gamma = k_z/k_0; \tag{15}$$

and k_0 is the wavenumber constant of the oblique wave. When $\nu_z = r$, the z-component of the wavenumber constant k_z becomes k_{zr} by introducing

$$k_{zr} = \frac{k_0 r L_z}{c} = \frac{n_r \pi}{L_z} \quad (\text{m}^{-1}),$$

then the equation

$$n_r = 2\left(\frac{L_z}{c}\right)^2 f_0 r \tag{16}$$

can be obtained, where f_0 is the frequency of the oblique wave. The average number of reflections at the z-walls, ν_{zr}, where the z-components of the wavenumber range from 0 to k_{zr}, is given by

$$\nu_{zr} = \tfrac{1}{2} r \quad (\text{s}^{-1}). \tag{17}$$

Reverberation Time in a Nearly Two-Dimensional Diffuse Field Suppose that there are n_r groups of nearly xy-tangential waves whose z-components of the wavenumber range from 0 to k_{zr}. Introducing the mean free path m_{xy} in the two-dimensional diffuse field, the average number of reflections at the x- and y-walls in the nearly two-dimensional field, ν_{xyr}, becomes

$$\nu_{xyr} = \frac{n_r c}{m_{xy}} \quad (\text{s}^{-1}). \tag{18}$$

The ratio of the average number of reflections at the z-walls to that of the other side walls (x- and y-walls), μ, becomes

$$\mu = \frac{r/2}{n_r c/m_{xy}} = \frac{m_{xy}c}{L_z^2 4 f_0}. \tag{19}$$

Following Eq. (13), the reverberation time in a nearly two-dimensional diffuse field is given by

$$T_{\text{Al}-xy} = \frac{0.128S_{2d}}{-\ln(1-\overline{\alpha}_{\text{Al}-xy})L_{2d}} \quad \text{(s)}, \tag{20}$$

where the averaged absorption coefficient $\overline{\alpha}_{\text{Al}-xy}$ is given by

$$\begin{aligned} \overline{\alpha}_{\text{Al}-xy} &= \overline{\alpha}_{xy}\left(1 - \frac{\nu_{zr}}{\nu_{xyr} + \nu_{zr}}\right) + \overline{\alpha}_z\left(\frac{\nu_{zr}}{\mu_{xyr} + \nu_{zr}}\right) \\ &\approx \overline{\alpha}_{xy}(1-\mu). \end{aligned} \tag{21}$$

Here $\overline{\alpha}_z$ is the averaged absorption coefficient of the z-wall. Figure 2 illustrates an example[6] of the frequency charactristic that follows Eq. (20). The frequency characteristic of the reverberation time at the midfrequency range is not due to the absorption material properties but to the geometric conditions of the sound field.

2.5 Power Law Decay Formula for a Diffuse Field

A nonexponential decay formula is derived even in a three-dimensional diffuse field.[8] In wave theory, many

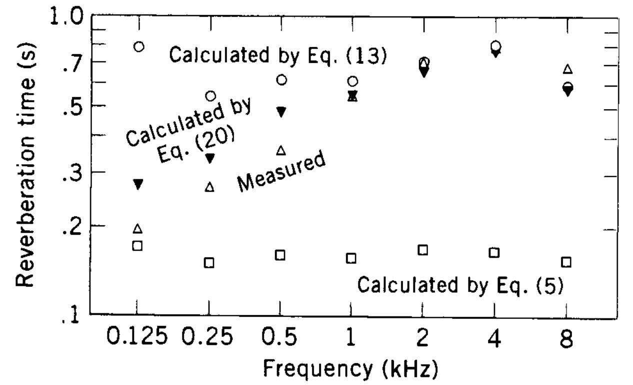


Fig. 2 Sample of frequency characteristics of reverberation time in a nearly two-dimensional diffuse field where the side walls perpendicular to the floor are hard and both floor and ceiling are soft. (From Ref. 6.)

wave modes are excited in a given frequency range. The energy decay is given by an average of exponential functions with different decay rates:

$$E(t) = E_0 \int_0^{\infty} P(\gamma) e^{-\gamma t}\, d\gamma \qquad (\mathrm{J/m^3}), \qquad (22)$$

where γ denotes a decay constant and $P(\gamma)$ is the distribution function for the decay rate.

Suppose that there is only one type of absorber on the surfaces of the room. The decay constant is proportional to the energy incident on the absorber. If we take an area S covered by an absorbing material, the incident energy flux can be expressed as

$$J_{\mathrm{in},S} = \int_S J_{\mathrm{in}}(r)\, dr \sim \sum_{n=1}^{N} p_n^2 \qquad (\mathrm{W}), \qquad (23)$$

where p_n^2 denotes the squared sound pressure data sampled at a point on the area S, N is the total number of "independent" samples (see Sections 4.2 and 4.3) in the data, $N \approx S/(\frac{1}{2}\lambda)^2$, and λ is the wavelength of the sound wave of interest.

As we see in Section 3.1, the square pressure data sampled in a diffuse field are distributed following an exponential distribution. Thus, the decay constant, written as a summation of N independent exponentially distributed variables, follows an Nth-order gamma distribution. Introducing this gamma distribution function into Eq. (23), we have a "power law" for the decay formula,

$$E(t) = E_0 \left(1 + \frac{\overline{\gamma} t}{N}\right)^{-N} \approx E_0 e^{-\overline{\gamma} t} \qquad (\mathrm{J/m^3}) \qquad (24)$$

as N becomes large, so that $\overline{\gamma} t/N \ll 1$, where $\overline{\gamma}$ denotes the average decay constant.[8]

3 TRANSFER FUNCTION IN A REVERBERATION FIELD

3.1 Transfer Function

The transfer function (TF) between a pair of source and receiving positions is obtained as the Fourier transform of the impulse response between the pair of positions. The TF in the space with a simple harmonic source at X_s and observer X_0 is given by a modal expansion form as[2]

$$H(\omega) = \mathrm{const} \sum_M \frac{\Psi_M(x_s)\Psi_M(x_0)}{\omega^2 - \Omega_M^2} \qquad (\mathrm{Pa \cdot s/m^3}), \qquad (25)$$

where Ψ_M are eigenfunctions (modal functions), $\omega = 2\pi f$, f is the frequency in hertz, and Ω_M are the eigenfrequencies.

3.2 Poles and Zeros of Transfer Functions

The TF can be rewritten as

$$H(\omega) = K \frac{(\omega - \omega_a)(\omega - \omega_b)\cdots}{(\omega - \omega_1)(\omega - \omega_2)\cdots} \qquad (\mathrm{Pa \cdot s/m^3}), \qquad (26)$$

where ω_1, ω_2 are the poles—the singularities of the transfer function—ω_a, ω_b are the zeros of the function, and K is a constant. The poles are independent of the location of the source or observation point. The zeros are dependent on the locations in the space.[9] Since an $\exp(j\omega t)$ time dependence is assumed, all the poles must be in the upper half-plane because of the stability or causality of the system (Fig. 3).

3.3 Modal Density and Modal Overlap of the TF

The number of eigenfrequencies in a unit frequency interval is called the modal density. The modal density is given by[10]

$$n_p(\omega) \approx \frac{V\omega^2}{2\pi^2 c^3} \qquad (\mathrm{s/rad}), \qquad (27)$$

where V is the volume (in cubic metres). Many modes (modal functions) are excited simultaneously even by a

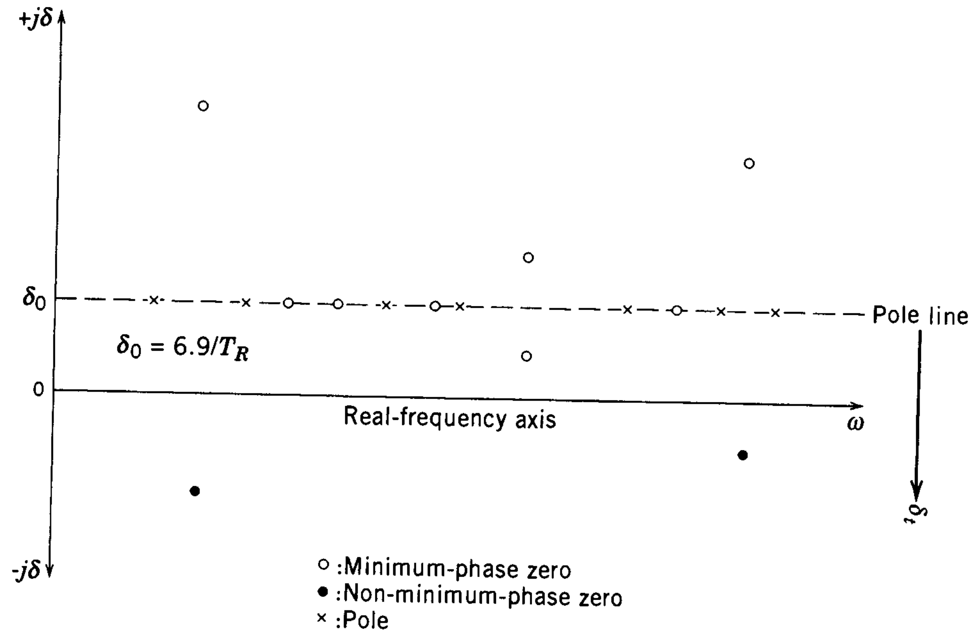


Fig. 3 Pole–zero pattern in the complex-frequency plane. (From Ref. 17.)

single frequency in a room. The number of such simultaneously excited modes is called the modal overlap,[10] which is defined as

$$M = Bn_p(\omega) = \pi\delta_0 n_p(\omega), \qquad (28a)$$

where B is the modal bandwidth defined by

$$B = \frac{1}{H^2_{0,\max}} \int_0^\infty |H_0(\omega)|^2 \, d\omega = \pi\delta_0 \qquad \text{(rad/s).} \quad (28b)$$

H_0 is the frequency response of a single mode, and $H_{0,\max}$ is the maximum of the absolute value of the response. We assume here that all the modal functions have the same shape.

3.4 Distribution of Eigenfrequencies

The poles are determined by the eigenfrequencies and lie along the pole line.[11] The eigenfrequencies of a rectangular room of dimensions (L_x, L_y, L_z) with rigid boundaries are given by

$$\omega_{lmn} = \frac{c\pi}{L} \sqrt{(al)^2 + (bm)^2 + (cn)^2} \qquad \text{(rad/s),} \quad (29)$$

where l, m, n are integer numbers; the lengths of the sides (in metres) are $L = aL_x = bL_y = cL_z$; and a, b, c are the ratios of the lengths. A Poisson model that results in an exponential distribution of the eigenfrequency spacings is not adequate to practical cases.[10] Groups of modes that have degenerate eigenfrequencies will suffer a splitting of the eigenfrequencies by the perturbations in practical systems.[12]

3.5 Distribution of Zeros

The poles are distributed along the pole line; however, the zeros are distributed two dimensionally throughout the complex-frequency plane.

Zeros in a Law Modal Overlap Condition The TF is approximately written as[13, 14] (using a dimensionless form)

$$H(\omega) \approx \frac{A}{\omega - \omega_1} + \frac{B}{\omega - \omega_2} + R. \qquad (30)$$

The primary part will dominate the TF in the modal expansion, with the rest of the terms producing a small remainder function R. A minimum-phase zero is produced between the two adjacent poles when they have residues of the same sign. Such a zero is located on the line connecting the poles (the pole line), as shown in Fig. 4. However, if the poles have residues of opposite signs, the occurrence of zeros will depend strongly on the remainder function R.[13, 14]

Two cases of zeros are produced from residues of opposite signs and the remainder function, as shown in Fig. 5. One is a double zero between two adjacent poles.

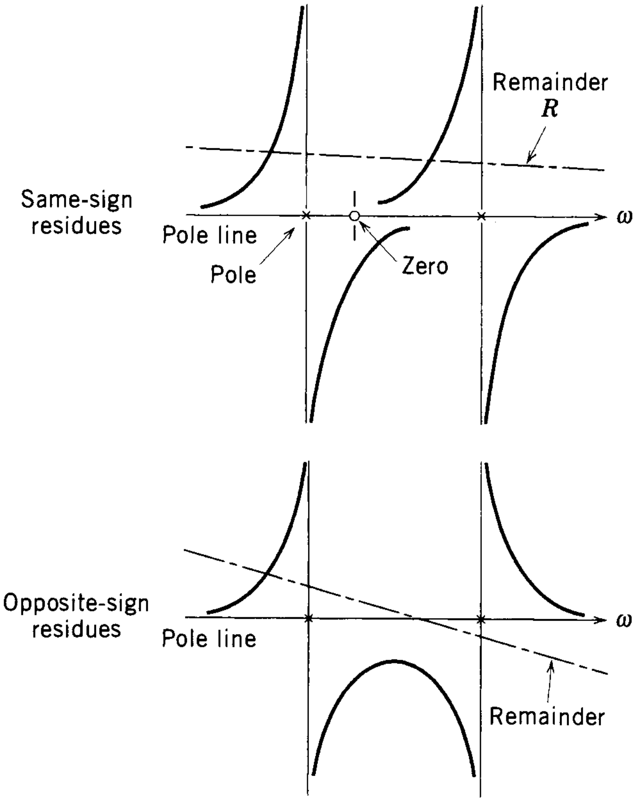


Fig. 4 Possibility of formation of a zero in the interval, depending on the relative signs of residues.

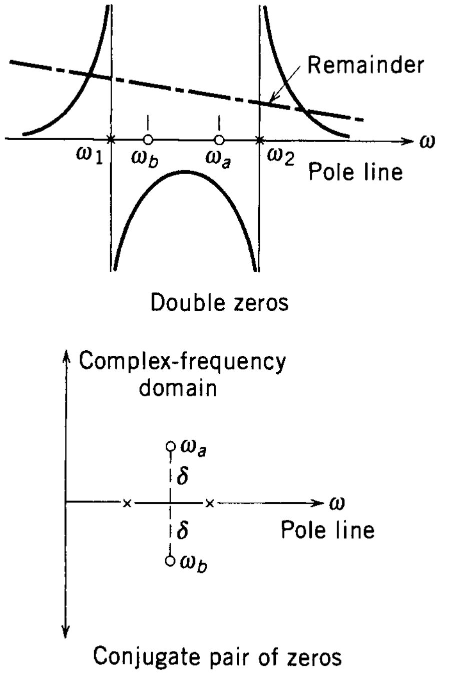


Fig. 5 Zeros from opposite-sign residues and remainder. (From Ref. 11.)

Both of these double zeros are minimum phase because they are located on the pole line. Another case produces a conjugate pair of zeros in the complex-frequency plane. Those zeros are located symmetrically with respect to each other at equal distances from the pole line.[11] One of these zeros can be non-minimum phase.

Zeros under a High Modal Overlap Condition

We take again a TF written as a modal expansion form by Eq. (25). When the source and observer are at the same location, all the residues are positive. Thus, the poles interlace with zeros, and the number of poles and zeros are equal. As source and receiver move apart, the zeros migrate (poles do not move). Some move above the line, an equal number move symmetrically below the line, and the remainder stay on the line.

If the modal functions are real, then, by symmetry,[11] the remaining zeros are evenly distributed above and below the pole line, which means that one-fourth of them are below the pole line. The possible number of zeros below the real frequency axis is $\frac{1}{4}N_p$, where $N_p = N_z^+ + N_z^-$, N_z^+, and N_z^- denote the number of non-minimum- and minimum-phase zeros, respectively. But this possible number of zeros is reduced as $\delta_0 = 6.9/T_R$ increases.

We will consider the TF for a reverberant field to be a random process in the complex-frequency domain.[15] Following the random process theory,[16] we postulate that the zeros are distributed around the pole line following a Cauchy distribution,[17]

$$p(\beta) = \frac{4/\pi}{1+4\beta^2}, \tag{31}$$

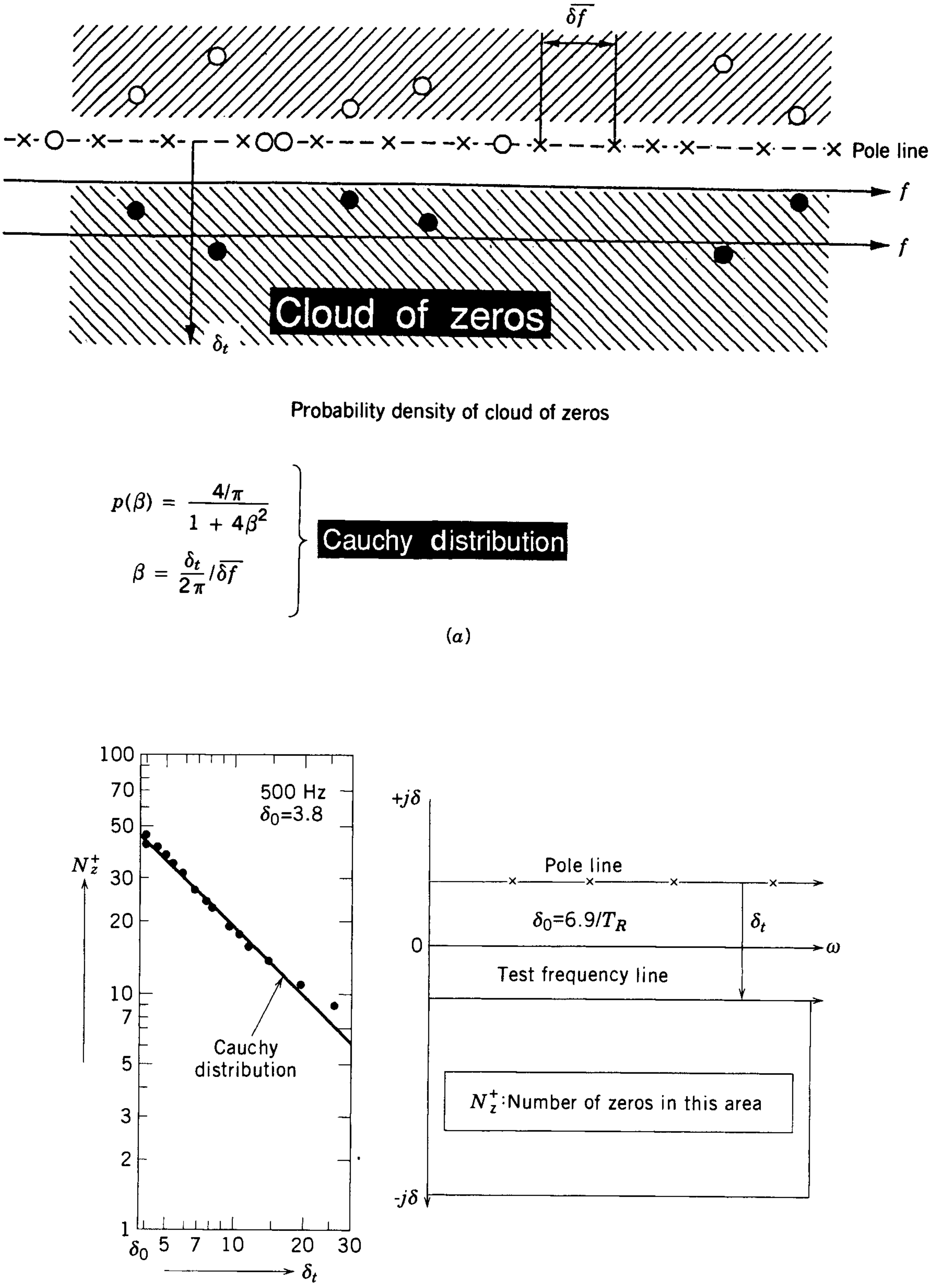
where $\beta = \delta_t/\overline{\delta\omega}$, $\overline{\delta\omega} = 1/n_p(\omega)$ being the average pole spacing (Fig. 6*a*).

The average number of zeros below the test frequency line in the lower half plane is illustrated in Fig. 6*b*[17] for four measured TFs. The result demonstrates that the distribution of zeros followed the solid line, which decreases following the Cauchy distribution by Eq. (31). For high modal overlap, the Cauchy distribution reduces to

$$p(\beta) = \frac{1}{\pi\beta^2}. \tag{32}$$

The number of non-minimum-phase zeros in a frequency band $\Delta\omega$ is

$$N_z^+ = \frac{n_p(\omega)\,\Delta\omega/4}{M} = \frac{\Delta\omega/4}{\pi\delta_t}. \tag{33}$$


$\overline{\delta f}$
Pole line
f
f
Cloud of zeros
δ_t
Probability density of cloud of zeros
$p(\beta) = \frac{4/\pi}{1 + 4\beta^2}$
$\beta = \frac{\delta_t}{2\pi}/\overline{\delta f}$
Cauchy distribution
100
70
50
N_z^+
30
20
10
7
5
3
2
1
500 Hz
δ_0=3.8
Cauchy distribution
δ_0
5
7
10
20
30
δ_t
$+j\delta$
Pole line
$\delta_0=6.9/T_R$
δ_t
0
ω
Test frequency line
N_z^+:Number of zeros in this area
$-j\delta$

(*a*)

(*b*)

Fig. 6 (*a*) Distribution of zeros of a transfer function. (*b*) Number of non-minimum-phase zeros in 500-Hz 1/1 octave band. (Part *b* from Ref. 17.)

The densities of the non-minimum-phase zeros is independent of the frequency, while the number of poles increases as the frequency becomes high.

3.6 Magnitude and Phase of Reverberation Transfer Functions

Distribution Function of Magnitudes The real and imaginary parts of the TF can be assumed to be mutually independent Gaussian variables.[15] The magnitude of the TF is distributed following a Rayleigh distribution[18]:

$$f(|H|=x)=\frac{2x}{\overline{|H|^2}}\exp\left(-\frac{x^2}{\overline{|H|^2}}\right), \qquad (34)$$

where $\overline{|H|^2}$ denotes the average of $|H|^2$.

Propagation Phase The growth of phase between source and observation locations in a one-dimensional system is simply visualized.[13,14] As shown in Fig. 7, suppose that at the nth eigenfrequency the x_0 and x_s are separated by p nodes so that $\mathrm{int}[(k/\pi)(x_0 - x_s)] = p$, where k is the wavenumber (in reciprocal metres). For higher eigenfrequencies, these locations stay in phase until p increases by unity. Then $\mathrm{int}[(k/\pi)(x_0 - x_s)] = p + 1$ and the phase between observation and source locations has increased by π. The phase trend in a one-dimensional case where the number of nodes increases in an orderly fashion between source and receiver is[13,14]

$$\Phi = k(x_0 - x_s) \quad \text{(rad)}. \qquad (35)$$

The accumulated phase of a one-dimensional system is just the same as the phase delay due to the direct wave propagation from source to observer. We call this the *propagation phase*.[13,14]

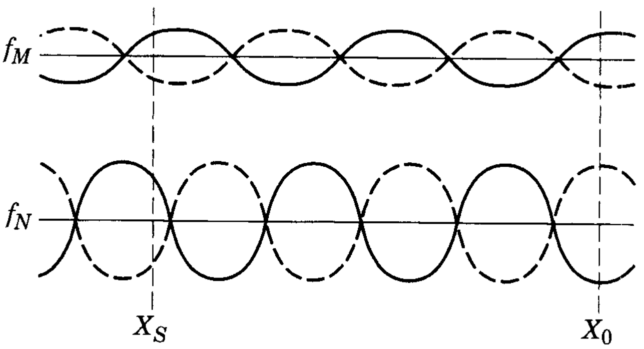


Fig. 7 Standing-wave pattern in a one-dimensional system showing phase advances as nodal pattern changes. (From Ref. 14.)

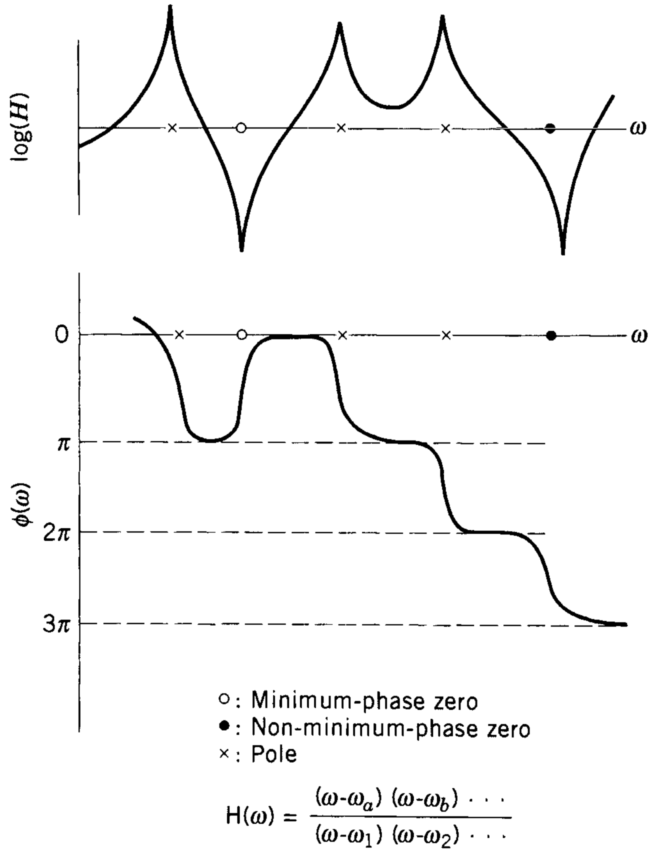

log(H)
ω
φ(ω)
0
π
2π
3π
ω
o: Minimum-phase zero
●: Non-minimum-phase zero
×: Pole
$H(\omega) = \frac{(\omega-\omega_a)(\omega-\omega_b)\cdots}{(\omega-\omega_1)(\omega-\omega_2)\cdots}$

Fig. 8 Poles, zeros, and accumulated phase. (From Ref. 17.)

Reverberation Phase The two- and three-dimensional cases are not as orderly as the one-dimensional case, however. The accumulated phase can be determined by both the poles and zeros (Fig. 8). If there are N_p poles, and if there are N_z^- and N_z^+ zeros in the upper and lower half planes, respectively, then the accumulated phase is

$$\Phi = \pi(N_p + N_z^+ - N_z^-) = 2\pi N_z^+ \quad \text{(rad)}, \qquad (36)$$

where $N_p = N_z^+ + N_z^-$. We can expect a large number of zeros to lie in the lower half plane; thus the accumulated phase by Eq. (36), which we call the *reverberation phase*, is much greater than the propagation phase in a one-dimensional system. Figure 9 illustrates a sample of the reverberation phase of the TF. We can see that the accumulated phase is approximately estimated by Eqs. (36) and (33) at various conditions of damping (δt).

3.7 Modulation Transfer Function

The modulation transfer function (MTF) was introduced as a measure in room acoustics for assessing the effect of

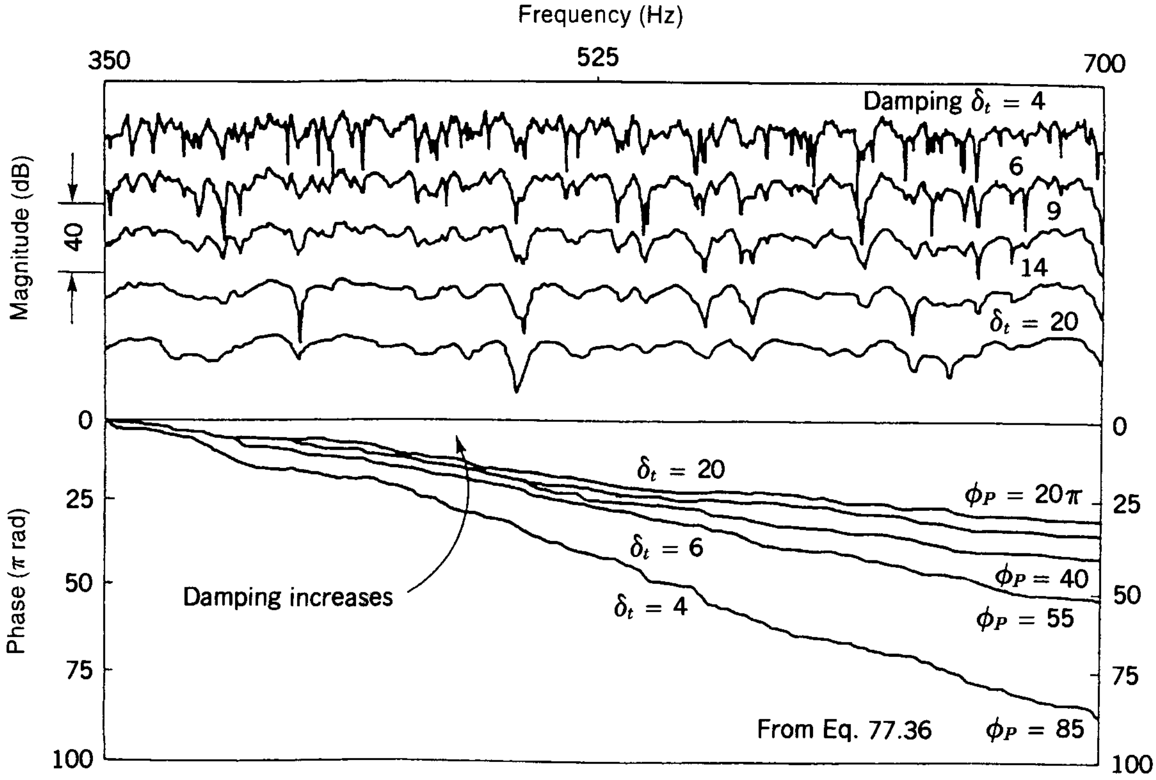


Fig. 9 Magnitude and phase of transfer functions. (From M. Tohyama, R. H. Lyon, and J. Koike, *J. Acoust. Soc. Am.*, Vol. 95, 1994, pp. 286–296.)

an enclosure on speech intelligibility.[19] A speech signal is nonstationary, but the temporal dynamic changes of speech signals are fairly slow; thus the temporal feature can be modeled by a modulated white noise. As shown in Fig. 10, for an input signal with a varying intensity of a white noise,

$$\bar{I}_i[1 + \cos 2\pi Ft],$$

with F the modulation frequency; the output signal is given as

$$\bar{I}_0[1 - m\cos 2\pi F(t - \tau)],$$

with m the modulation index and τ the time lag due to the transmission. We define the function $m(F)$ as the MTF. The MTF is only of interest for the range of F relevant for speech. When the entire relevant range, from 0.4 Hz up to 20 Hz, is transmitted faithfully, intelligibility will be excellent.

When considering only the influence of reverberation, the MTF is the Fourier transform of the squared impulse response.[20] The input intensity is rewritten as

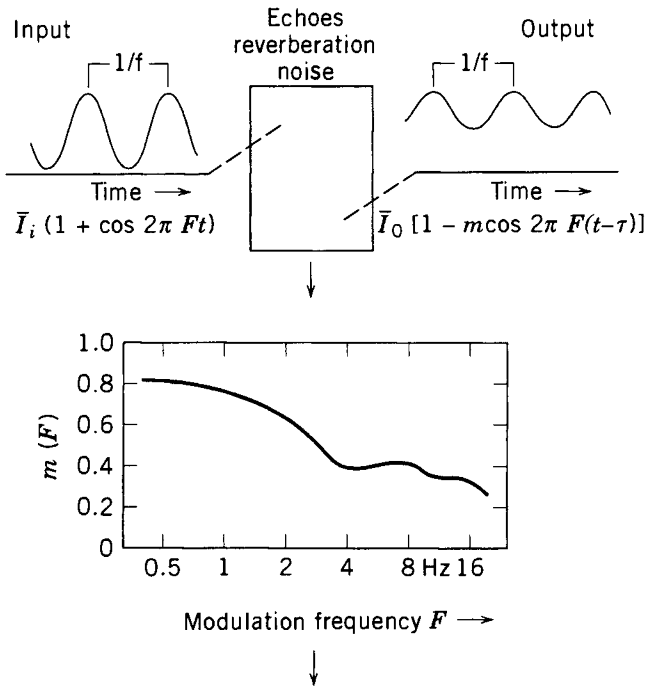

Input
1/f
Time →
$\bar{I}_i\ (1 + \cos 2\pi\ Ft)$
Echoes reverberation noise
Output
1/f
Time →
$\bar{I}_0\ [1 - m\cos 2\pi\ F(t-\tau)]$
1.0
0.8
0.6
0.4
0.2
0
$m\ (F)$
0.5
1
2
4
8 Hz 16
Modulation frequency F →

Fig. 10 Schematic illustration of the modulation transfer function. (From Ref. 19.)

$$I_i(t) = \bar{I}_i[1 + \mathrm{Re}(e^{j2\pi Ft})] \qquad (\mathrm{W/m^2}). \qquad (37)$$

The output intensity is given by

$$I_0(t) = \int_0^\infty I_i(t - t')h^2(t')\,dt' = aI_i\left[1 + \frac{1}{a}\,\mathrm{Re}(be^{j2\pi Ft})\right]$$
$$(\mathrm{W/m^2}), \qquad (38)$$

where

$$a = \int_0^\infty h^2(t')\,dt', \qquad b = \int_0^\infty e^{-j2\pi Ft'}h^2(t')\,dt'. \qquad (39)$$

The MTF is obtained as

$$m(F) = \frac{\left|\int_0^\infty e^{-j2\pi Ft'}h^2(t')\,dt'\right|}{\int_0^\infty h^2(t')\,dt'}. \qquad (40)$$

The Fourier transform of the squared impulse response equals the complex autocorrelation function of the TF.[15]

4 SPATIAL DISTRIBUTION OF REVERBERATION SOUNDS

4.1 Potential Energy Distribution of Reverberation Sounds

There are no regular patterns of loops and nodes in a reverberant space because of the high modal overlap. Suppose the mean-squared pressure (proportional to the potential energy density) is measured at different points in the room. The sound pressure at a point is

$$p(t) = p_0 \sum_{i=1}^{n} \cos(\omega t + \alpha_i). \qquad (41)$$

At this point, there is a set of n phase angles α_i distributed as $0, \ldots, 2\pi$ and a fixed value of mean-squared pressure at the point,

$$\overline{p_1^2} = \frac{1}{2}\,p_0^2\left[\left(\sum_{i=1}^{n} \cos\alpha_i\right)^2 + \left(\sum_{i=1}^{n} \sin\alpha_i\right)^2\right]$$
$$= \frac{1}{2}\,p_0^2(r_1^2 + s_1^2) \qquad (\mathrm{Pa^2}). \qquad (42)$$

Sampling at many well-spaced points gives the distribution of values of $\overline{p_k^2}$.

When α_i takes on any value between 0 and 2π, the quantities $\cos\alpha_i$ are distributed between the limits ±1. The distribution of $\sum_{i=1}^{n} \cos\alpha_i = r_k$ is not normal, but for a large number of n, we can treat r_k and s_k as normally distributed variables. The mean-squared pressure is a random variable given by the squared sum of two normally distributed variables which have zero mean and variance $\frac{1}{2}n$ and are mutually independent.[18,21] Normalizing so that the expected value of $\overline{p_k^2}$ (ensemble average of $\overline{p_k^2} = \langle\overline{p_k^2}\rangle$) is unity, which corresponds to putting

$$\langle\overline{p_k^2}\rangle = \tfrac{1}{2}np_0^2 = 1 \qquad (\mathrm{Pa^2}), \qquad (43)$$

the distribution of the sampled value of $\overline{p_i^2} = \overline{p^2}$ has the probability density function[18,21]

$$P(\overline{p^2} = x) = e^{-x}. \qquad (44)$$

The probability density function of $\overline{p^2}$ follows an exponential curve independent of exciting frequency. Even in a large reverberation room, we can no longer expect a uniform distribution of squared sound pressure under the condition that a sound source radiates a pure sinusoidal wave. The exponential distribution belongs to a family of gamma distributions.

4.2 Spatial Correlations for Reverberation Sounds

Suppose that many observations are taken with separation r between two points of r_i and r_j, where $r = |r_i - r_j|$ in a reverberation field. Since the plane waves are uniformly distributed in all directions in the ensemble, the spatial cross correlation of the sound pressure can be written as an ensemble average,[18,22]

$$\langle\overline{p(r_i,t)p(r_j,t)}\rangle = \frac{1}{4}\int_0^\pi \cos\{kr\cos\theta\}\sin\theta\,d\theta$$
$$= \frac{1}{2}\,\frac{\sin kr}{kr}, \qquad (45a)$$

where the overbar denotes the time average. Consequently, the normalized spatial cross correlation is

$$R_0(r) = \frac{\sin kr}{kr}. \qquad (45b)$$

4.3 Spatial Correlations for Squared Reverberant Sounds

To ensure independent sampling $\overline{p^2}$, it is necessary that spatial correlation between paired samples be negligibly small. Similar to the discussions above, taking the ensemble average of the paired mean-square pressure data, we have the normalized correlation coefficient for the mean-squared pressure data sampled at two locations whose separation is r^{22}:

$$R_0(r) = \left(\frac{\sin kr}{kr}\right)^2. \qquad (46)$$

Sound fields are described deterministically in principle based on the wave equation. But there are many parameters in sound fields, most of which are quite difficult to be precisely controlled and formulated, such as room temperature, boundary conditions, and sound source characteristics. The statistical approach is a practical way of predicting the sound field characteristics in rooms.

REFERENCES

1. C. W. Kosten, "The Mean Free Path in Room Acoustics," *Acustica*, Vol. 10, 1960, pp. 245–250.
2. H. Kuttruff, *Room Acoustics*, 3rd ed., Elsevier Science, New York 1991, pp. 114–120.
3. M. R. Schroeder, and D. Hackman, "Iterative Calculation of Reverberation Time," *Acustica*, Vol. 45, 1980, pp. 269–273.
4. M. Barron, "Growth and Decay of Sound Intensity in Rooms According to Some Formulae of Geometric Acoustics Theory," *J. Sound Vib.*, Vol. 27, 1973, pp. 183–196.
5. Y. Hirata, "Dependence of the Curvature of Sound Decay Curves and Absorption Distribution on Room Shapes," *J. Sound Vib.*, Vol. 84, 1982, pp. 509–517.
6. M. Tohyama and A. Suzuki, "Reverberation Time in an Almost-Two-Dimensional Diffuse Field," *J. Sound Vib.*, Vol. 111, 1986, pp. 391–398.
7. L. Batchelder, "Reciprocal of the Mean Free Path," *J. Acoust. Soc. Am.*, Vol. 36, 1964, pp. 551–555.
8. F. Kawakami and K. Yamaguchi, "A Systematic Study of Power-Law Decays in Reverberation Rooms," *J. Acoust. Soc. Am.*, Vol. 80, 1986, pp. 543–554.
9. R. H. Lyon, *Machinery Noise and Diagnostics*, Butterworth, Boston and London, 1987.
10. R. H. Lyon, "Statistical Analysis of Power Injection and Response in Structures and Rooms," *J. Acoust. Soc. Am.*, Vol. 45, 1969, pp. 545–565.
11. M. Tohyama and R. H. Lyon, "Zeros of a Transfer Function in a Multi-Degree-of-Freedom Vibrating System," *J. Acoust. Soc. Am.*, Vol. 86, 1989, pp. 1854–1863.
12. J. L. Davy, "The Distribution of Modal Frequencies in a Reverberation Room," *Proc. Inter-Noise '90*, 1990, pp. 159–164.
13. R. H. Lyon, "Progressive Phase Trends in Multi-Degree-of-Freedom Systems," *J. Acoust. Soc. Am.*, Vol. 73, 1983, pp. 1223–1228.
14. R. H. Lyon, "Range and Frequency Dependence of Transfer Function Phase," *J. Acoust. Soc. Am.*, Vol. 76, 1984, pp. 1435–1437.
15. M. R. Schroeder, "Frequency-Correlation Functions of Frequency Responses in Rooms," *J. Acoust. Soc. Am.*, Vol. 34, 1962, pp. 1819–1823.
16. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed., McGraw-Hill, New York, 1991.
17. M. Tohyama, R. H. Lyon, and T. Koike, "Reverberant Phase in a Room and Zeros in the Complex Frequency Plane," *J. Acoust. Soc. Am.*, Vol. 89, 1991, pp. 1701–1707.
18. K. J. Ebeling, "Statistical Properties of Random Wave Fields," in W. P. Mason, and R. N. Thurston (Eds.), *Physical Acoustics*, Academic, New York, 1984, pp. 223–310.
19. T. Houtgast, H. J. M. Steeneken, and R. Plomp, "Predicting Speech Intelligibility in Rooms from the Modulation Transfer Function. I. General Room Acoustics," *Acustica*, Vol. 46, 1980, pp. 60–72.
20. M. R. Schroeder, "Modulation Transfer Functions: Definition and Measurement," *Acustica*, Vol. 49, 1981, pp. 179–182.
21. R. V. Waterhouse, "Statistical Properties of Reverberant Sound Fields," *J. Acoust. Soc. Am.*, Vol. 43, 1968, pp. 1436–1444.
22. W. T. Chu, "Note on the Independent Sampling of Mean-Square Pressure in Reverberant Sound Fields," *J. Acoust. Soc. Am.*, Vol. 72, 1982, pp. 196–199.

78

STATISTICAL MODELING OF VIBRATING SYSTEMS

JEROME E. MANNING

1 INTRODUCTION

In statistical modeling, the properties of the vibrating system are assumed to be drawn from a random distribution. This allows great simplification in the analysis, whereby modal densities, average mobility functions, and statistical energy analysis (SEA) can be used to obtain response estimates and transfer functions. To demonstrate the benefits of a statistical approach, the point conductance is derived using a modal analysis, with the resonance frequencies and mode shapes assumed to be random variables. The average conductance is found to be a simple function of the modal density. A graphical representation of the modes in wavenumber space is used to calculate the modal density. The use of finite-element models to determine the modal density is also discussed.

Statistical energy analysis combines statistical modeling with an energy flow formulation. Using SEA, the energy flow between coupled subsystems is defined in terms of the average modal energies of the two subsystems and a coupling factor. Equations of motion are obtained by equating the time-average power input to each subsystem with the sum of the power dissipated and the power transmitted to other subsystems. Techniques to determine the coupling and dissipation factors needed for the SEA analysis are discussed.

2 STATISTICAL APPROACH TO DYNAMIC ANALYSIS

Earlier chapters have presented a variety of techniques to study the dynamic response of complex acoustic and structural systems. By and large, these techniques have used a deterministic approach. In the analytical techniques, it has been assumed that the system being studied can be defined in terms of an idealized mathematical model. Techniques based on the use of measured data have assumed that the underlying physical properties of the system are well defined and time invariant. Although the excitation of the system was considered to be a random process in several of the earlier chapters, the concept that the system itself has random properties was not pursued.

The most obvious sources of randomness are manufacturing tolerances and material property variations. Although variations in geometry may be small and have a negligible effect on the low-frequency dynamics of the system, their effect at higher frequencies cannot be neglected. Another source of randomness is uncertainty in the definition of the parameters needed to define a deterministic model. For example, the geometry and boundary conditions of a room will change with the arrangement of furnishings and partitions. Vehicles will also change due to different configurations of equipment and loadings. Finally, during the early phases of design, the details of the product or building being designed are not always well defined. This makes it necessary for the analyst to make an intelligent guess for the values of certain parameters. If these parameters do have a major effect on the response being predicted, the consequences of a poor guess are not serious. On the other hand, if the parameters do have a major effect, the consequences of a poor guess can be catastrophic. Although deterministic methods of analysis have the potential to give exact solutions, they often do not, because of errors and uncertainties in the definition of the required parameters.

In Chapter 63, the vibratory response *of a dynamic* system was represented in terms of the modes of vibration of the system. Both theoretical and experimental

---

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker
Copyright © 1997 John Wiley & Sons, Inc.

techniques to obtain the required mode shapes and resonance frequencies were described. A statistical approach will now be followed in which resonance frequencies and mode shapes are assumed to be random variables. This approach will result in great simplifications where average mobility functions and power inputs from excitation sources can be defined simply in terms of the modal density and structural mass or acoustic compliance of the system. Modal densities in turn can be expressed in terms of the dimensions of the system and a dispersion relation between wavenumber and frequency.

If the properties describing a dynamic system can be accurately defined, the modes of the system can be found mathematically in terms of the eigenvalues and eigenfunctions of a mathematical model of the system, or they can be found experimentally as the resonance frequencies and response amplitudes obtained during a modal test. The availability of finite-element models and computer software makes it possible today to determine the modes of very complex and large structures. Although it might be argued that the computational cost to determine a sufficient number of modes to analyze acoustic and structural systems at high frequencies is too high, that cost is dropping each year with the development of faster and less expensive computers. Thus, many vibration analysts believe that analysis based on modes determined from a finite-element model can provide them with the answers they need. Similarly, many engineers who are more inclined toward the use of measured data believe that analysis based on modes determined from a modal test will suffice.

The accuracy of a modal analysis depends on the accuracy with which the modal parameters can be determined and on the number of modes that are included in the analysis. Since the modes form a mathematically complete set of functions, a high degree of accuracy can be obtained by including a large number of modes in the analysis. In practice, the accuracy with which the model parameters can be determined decreases for the higher order modes. Thus, the accuracy of the modal analysis depends to a large extent on the number of modes needed to describe the vibration of the system. At low frequencies, for lightly damped systems, the response can be accurately described in terms of the response of a few modes. On the other hand, at high frequencies and for highly damped systems, a large number of modes are required, and the accuracy of the modal analysis decreases. In this case, a statistical description of the system is warranted. Using a statistical description avoids the need for an accurate determination of the resonance frequencies and mode shapes, thereby eliminating the main source of error in the modal analysis. Of course, in using a statistical description, the ability to determine the exact response at specific locations and frequencies is lost. Instead, a statistical description of the response is obtained.

2.1 Mobility Formulation

To illustrate the use of analysis and the application of a statistical approach, the response of a linear system to a point excitation with harmonic $e^{j\omega t}$ time dependence is considered. For a structural system, the ratio of the complex amplitude of the response velocity to the complex amplitude of the excitation force is defined as the point mobility for the system. A transfer mobility is used if the location of the response point is different than that of the excitation point. A drive point mobility is used if the locations of the response and drive points are the same. The drive point mobility at coordinate location x can be written as a summation of modal responses,

$$Y_{\text{pt}}(x,\omega) = \sum_{i=1}^{\infty} \frac{1}{M_i} \frac{j\omega\psi_i^2(x)}{(\omega_i^2 - \omega^2) + j\omega\omega_i\eta_i}, \tag{1}$$

where $Y_{\text{pt}}(x,\omega)$ is the drive point mobility for the structure, M_i is the modal mass, $\psi_i(x)$ is the mode shape for the ith mode, ω is the radian frequency of excitation, ω_i is the resonance frequency of the ith mode, η_i is the damping loss factor, $j = \sqrt{-1}$, and the summation is over all modes of the system. The mobility consists of a real part (the conductance) and an imaginary part (the susceptance). The conductance is often of greater interest, since the product of the conductance and the mean-square force is the input power to the system. The conductance and susceptance for a typical lightly damped system are shown in Fig. 1. For light damping, the conductance as a function of frequency shows a large peak at each resonance frequency. The amplitude of the peak is governed by the damping and by the value of the mode shape at the drive point.

Frequency Averages A frequency-averaged conductance is found by integrating the real part of the mobility over frequency and dividing by the bandwidth,

$$\langle G_{\text{pt}}(x,\omega)\rangle_{\Delta\omega} = \frac{1}{\Delta\omega}\int_{\Delta\omega} \text{Re}\{Y_{\text{pt}}(x,\omega)\}\Delta\omega, \tag{2}$$

where $\langle\cdot\rangle$ signifies an average, $\text{Re}\{\cdot\}$ signifies the real part of a complex number, and the integral is over the frequency band, $\Delta\omega$. For any particular frequency band, the average conductance is largely determined by the number of modes with resonance frequencies within the band. For light damping, the contribution to the integral

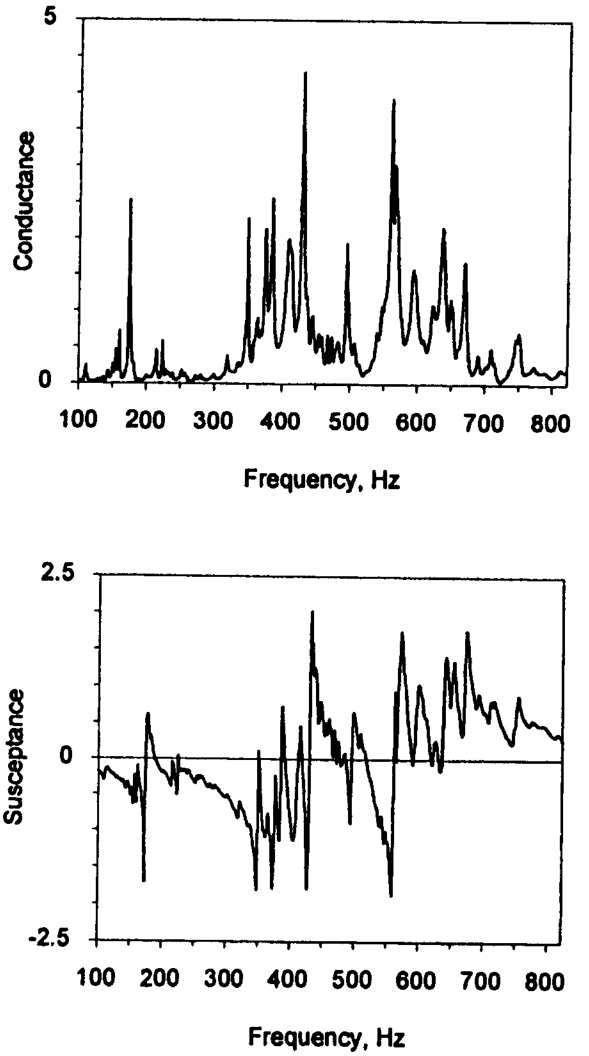


Fig. 1 Drive point conductance and susceptance for a typical structure.

in Eq. (2) from a single mode with a resonance frequency within the band is approximately $\psi_i^2\pi/2M_i$. If the resonance frequency of the mode is outside the band, $\Delta\omega$, the contribution to the integral is very small and can be ignored. Thus, the frequency-averaged conductance is given by

$$\langle G_{\text{pt}}(x,\omega)\rangle_{\Delta\omega} = \frac{1}{\Delta\omega}\,\frac{\pi}{2}\sum_{\substack{i\\ \text{modes}\\ \text{in}\,\Delta\omega}} \frac{\psi_i^2(x)}{M_i}, \tag{3}$$

where the summation is over all modes with resonance frequencies in the band. Note that the average conductance does not depend on the precise values for the resonance frequencies or the damping but only on the number of modes within the band and their mode shapes at the drive point.

Spatial Averages The statistical approach can be extended one step further by averaging the conductance over the spatial extent of the system. For a homogeneous system, the spatial-average value of the mode shape squared is equal to the modal mass divided by the physical mass of the system, M_i/M. Thus, the drive point conductance averaged both over a band of frequencies and over the spatial extent of the system is given by

$$\langle G_{\text{pt}}(x,\omega)\rangle_{\Delta\omega,x} = \frac{\pi}{2M}\,\frac{N_{\Delta\omega}}{\Delta\omega}, \tag{4}$$

where $N_{\Delta\omega}$ is the number of modes with resonance frequencies in the band $\Delta\omega$ (the "resonant mode count") and M is the physical mass of the structure. Equation (4) applies only to homogeneous structures. However, it can also be used for the general case if we replace M in Eq. (4) by the dynamic mass M_d, where

$$M_d(\omega) = \frac{1}{N_{\Delta\omega}}\sum_{\substack{i\\ \text{modes}\\ \text{in}\,\Delta\omega}} \frac{\langle\psi_i^2(x)\rangle_x}{M_i}. \tag{5}$$

The definition of a dynamic mass allows the expression given in Eq. (4) to be used for the average conductance of both homogeneous structures and nonhomogeneous structures, such as framed plates and structures loaded with components.

Acoustic Systems The formulation above is for a structural system in which the equations of motion are formulated in terms of a response variable such as displacement. For an acoustic system, the equations of motion are typically formulated in terms of pressure, a stress variable. In this case, a similar formulation can be carried out. However, the acoustic resistance (pressure divided by volume velocity) is obtained rather than the conductance, and the structural mass is replaced by the bulk compliance of the acoustic space,

$$\langle R_{\text{pt}}^A(x,\omega)\rangle_{\Delta\omega,x} = \frac{\pi}{2C_a}\,\frac{N_{\Delta\omega}}{\Delta\omega}, \tag{6}$$

where R_{pt}^A is the point acoustic resistance for the acoustic space and C_a is the bulk compliance of the space ($V/\rho c^2$ for a space with rigid walls).

Ensemble Averages An average mobility can also be obtained by averaging over an ensemble of structures with random parameters. This type of average has the advantage that it can be useful both for random excitation and for single-frequency excitation. In this case, an average is taken out over bands of frequency but over an ensemble of systems with randomly varying resonance frequencies and mode shapes. The average conductance for this ensemble of systems is given by

$$\langle G_{\mathrm{pt}}(x,\omega)\rangle_{\Delta\omega,x} = \frac{\pi}{2}\frac{n(\omega)}{M_d}, \qquad (7)$$

where $n(\omega)$ is the modal density for the system (the ensemble-average number of modes per unit radian frequency).

The frequency-average and ensemble-average conductances are approximately equal, since the product of the modal density and the bandwidth gives an estimate for the ensemble-average resonant mode count. The equality of the frequency average and ensemble average is analogous to the ergodic hypothesis in random process theory.

The susceptance or imaginary part of the mobility can also be determined from a modal summation. However, since the imaginary part of each term in the summation exhibits both positive and negative values, the number of terms that must be included in the summation to obtain a good estimate of the susceptance is much greater than required to obtain a good estimate of the average conductance.

2.2 Modal Density

The previous section shows how the use of a statistical description of a dynamic system can lead to a great simplification in the determination of the average structural conductance or acoustic resistance. The exact resonance frequency and mode shape for each individual mode are no longer needed. Instead, the resonant mode count or modal density can be used. The distinction between these two variables is often quite subtle. The term *mode count* is used to describe the number of modes with resonance frequencies within a given frequency band. The modal density, on the other hand, is a mathematical quantity that gives a statistical estimate of the average number of resonant modes per unit frequency for an ensemble of systems. If we define an ensemble of systems for which the geometry and material parameters vary randomly within manufacturing tolerances or design limits, the mode count will vary from system to system within the ensemble. The average mode count over the ensemble can be estimated from the modal density,

$$N_{\Delta\omega} = \int_{\omega_1}^{\omega_2} n(\omega)\, d\omega, \qquad (8)$$

where ω_1 and ω_2 are the lower and upper frequencies of the band, $\Delta\omega$. In many cases the modal density will be a fairly smooth function of frequency. The average mode count can then be obtained simply as the product of the modal density at the band center frequency and the frequency bandwidth.

Asymptotic Modal Densities In most cases, the mode count or modal density can be determined analytically using asymptotic modal densities that are valid at high frequencies, where many resonant modes exist even in narrow bands of frequency. The general use of these asymptotic modal densities to obtain the mode count has led to the idea that statistical modeling can only be used at high frequencies. This is indeed a misconception. As long as correct values are used for the modal density and the dynamic mass, the statistical modeling can be extended to low frequencies, where the number of modes is small. However, at these low frequencies the variance of the estimates may become quite large, so that the average value is not a good estimate for any individual structure.

One-Dimensional Systems The modes of a one-dimensional system have mode shapes that are functions of a single spatial coordinate. For a system with uniform properties, the mode shapes at interior positions away from any boundary are in the form of a sinusoid,

$$\psi_i(x) = A_i \sin(k_i x + \varphi_i), \qquad (9)$$

where k_i is a wavenumber describing the rate of variation of the mode shape with the spatial coordinate x and φ_i is a spatial phase factor that depends on boundary conditions. The wavenumber for a mode is related to the number of half-wavelengths within the length of the system,

$$k_i = \frac{\pi}{L}(i+\delta), \qquad (10)$$

where L is the length of the system, i is an integer, and δ is a small constant ($-\frac{1}{2} \le \delta \le \frac{1}{2}$) whose value depends on the boundary conditions at each end of the system.

The modes can be represented graphically in wavenumber space. For a one-dimensional system, wavenumber space is a single axis. Each mode can be represented by a point along the axis at the value k_i, as shown in Fig. 2. If the value of δ is constant from mode to mode, as

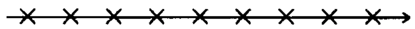


Fig. 2 Wavenumber space for a one-dimensional system.

would be the case for clamped or free boundary conditions, the spacing between modes in wavenumber space, δk, is simply the ratio π/L. If, on the other hand, the value of δ is random due, for example, to a random boundary condition, the average spacing between modes is also π/L. Although the exact position of a mode in wavenumber space depends on the value of δ, each mode will "occupy" a length of the wavenumber axis defined by the average spacing δk.

To determine the modal density, a relationship between wavenumber and frequency is needed. This relationship is the dispersion relation or characteristic equation for the system. For a simple one-dimensional acoustic duct the dispersion relation is $k = \omega/c$, where c is the speed of sound. For bending deformations of a one-dimensional beam (without shear deformations or rotational inertia) the dispersion relation is $k^4 = \omega^2 m/EI$, where m is the mass per unit length, E is Young's modulus for the material, and I is the bending moment of inertia. The dispersion relation allows mapping of the frequency range $\Delta\omega$ to the corresponding wavenumber range Δk. The average value of the mode count is then obtained by dividing Δk by δk. The modal density can now be written as the limit,

$$n(\omega) = \lim_{\Delta\omega \to 0} \frac{\Delta k}{\Delta\omega} \frac{1}{\delta k}. \tag{11}$$

The ratio $\Delta k/\Delta\omega$, in the limit as the frequency range approaches zero, becomes the derivative of the wavenumber with respect to frequency. This derivative is the inverse of $d\omega/dk$, which is the group speed for the system. Thus, the modal density for the general one-dimensional system becomes

$$n(\omega) = \frac{L}{\pi c_g}. \tag{12}$$

The modal density of the general one-dimensional system is simply related to the length of the system and the group speed. In this asymptotic result, which becomes exact as the number of modes increases, boundary conditions are not important. For the lower order modes, it is possible to correct the modal density for specific types of boundary conditions. However, the increase in accuracy may not be worth the effort.

The general result shows that the modal density depends on the group speed. Thus, the modal density of a one-dimensional acoustic lined duct should take into account the effect of the lining on the group speed. If the acoustic media is a fluid, the wall compliance can have a significant effect on the group speed and should be considered.

Two-Dimensional Systems The modes of a two-dimensional system have mode shapes that vary along two spatial coordinates. Like the one-dimensional system, the modes of the two-dimensional system can be represented graphically in wavenumber space. However, wavenumber space becomes a plane defined by two axes. Each mode can be represented by a point on the plane, so that the modes of the system can be represented by a lattice of points, as shown in Fig. 3. For a rectangular system, the lattice forms a rectangular grid of points. The average spacing between the points along the k_x axis is π/L_x, and the spacing between the points along the k_y axis is π/L_y, where L_x and L_y are the dimensions of the system. Thus, a mode occupies a small area in wavenumber space equal to π^2/A, where A is the area of the panel.

The dispersion relation for a two-dimensional system relates the frequency to the two wavenumber components k_x and k_y. For a simple rectangular acoustic layer, the dispersion relation is $k_x^2 + k_y^2 = \omega^2/c^2$, where c is the speed of sound. For bending deformations of a rectangular plate (without shear deformations or rotational inertia) the dispersion relation is $(k_x^2 + k_y^2)^2 = \omega^2 m/EI'$, where m is the mass per unit area and I' is the bending moment of inertia of the plate. The dispersion relation allows lines of constant frequency to be drawn in wavenumber space for the upper and lower frequencies of the band, $\Delta\omega$. For the simple systems above, these lines are quarter-circles forming an annular region, as shown in Fig. 3. The average value of the mode count is obtained by dividing the area of this region by the area occupied by a single mode. The modal density can now be written as the limit,

$$n(\omega) = \lim_{\Delta\omega \to 0} \frac{\Delta k}{\Delta\omega} \frac{kA}{2\pi}, \tag{13}$$

where $k^2 = k_x^2 + k_y^2$. The modal density for the general

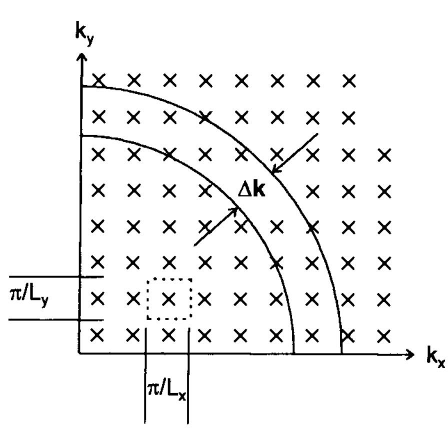


Fig. 3 Wavenumber space for a two-dimensional system.

two-dimensional system becomes

$$n(\omega) = \frac{kA}{2\pi c_g}. \tag{14}$$

Both the wavenumber and the group speed can be found from the dispersion relation.

The formulation of modal density using wavenumber space allows extension of the results to more complicated systems. For example, the modal density for bending modes of a fluid-loaded plate can be obtained by adjusting the group speed to account for the fluid loading. Similarly, the modal density for bending modes, including the effects of transverse shear deformations and rotary inertia, can be obtained by adjusting the group speed to account for these effects. The modal densities for cylindrical shells and orthotropic plates can be obtained using wavenumber space. For these cases, however, the lines of constant frequency will not be circular.

Three-Dimensional Systems The modes of a three-dimensional system have mode shapes that vary along three spatial coordinates. In this case, wavenumber space becomes a volume defined by three axes. Like the one-dimensional and two-dimensional systems, each mode can be represented by a point in wavenumber space, so that the modes of the system form a lattice of points in three dimensions. Each mode occupies a small volume in wavenumber space equal to π^3/V, where V is the volume of the system. The exact location of the mode within this volume will depend on the exact boundary conditions.

The dispersion relation for the three-dimensional system gives the frequency as a function of three wavenumbers. For a uniform acoustic space the dispersion relation is

$$\omega = c\sqrt{k_x^2 + k_y^2 + k_z^2}, \tag{15}$$

where c is the speed of sound. The dispersion relation allows two-dimensional surfaces of constant frequency to be drawn in wavenumber space. The average cumulative mode count can now be obtained by dividing the total volume under these surfaces by the average volume occupied by a single mode. The dispersion relation for the uniform acoustic space results in spherical surfaces with a volume of $\pi k^2/2$ (one-eighth sphere with radius k). Each mode occupies a volume equal to π^3/V, so that the cumulative mode count can be written as

$$N(\omega) = \frac{k^3 V}{6\pi^2}. \tag{16}$$

The modal density is found from the derivative of the cumulative mode count,

$$n(\omega) = \frac{k^2 V}{2\pi^2 c_g}. \tag{17}$$

The asymptotic modal density of the general three-dimensional space gives an accurate estimate of the mode count at high frequencies where many modes occur. It can also be used at lower frequencies as the ensemble average for systems with random boundary conditions. In room acoustics, the walls are commonly assumed to be rigid. In this case corrections to the asymptotic modal density can be introduced that improve the accuracy of the mode count at low frequencies. However, these corrections are only valid for a space in which the assumption of rigid walls is valid. Such an assumption would not be valid, for example, for a fluid-filled tank.

Finite-Element Modeling Finite-element models show great potential as a means to determine the correct mode count for complex structures and acoustic spaces at lower frequencies. Although these models may lose some accuracy in defining the resonance frequencies for the higher order modes, their use to determine the number of modes within defined frequency bands is justified. Two procedures can be used for determining a mode count from a finite-element model. In the first an eigenvalue analysis is used to obtain the resonance frequencies. The mode count is obtained by dividing the frequency range of interest into bands and counting the number of resonance frequencies in each band. In the second technique, the resonance frequencies are used to determine the frequency spacing between modes. The average frequency spacing is found by averaging over a set number of spacing intervals rather than over a set frequency band. Finally, the modal density is obtained from the inverse of the average spacing. Although the first technique is commonly used, the second technique is preferred, since it provides an estimate of the average modal density with a constant statistical accuracy.

3 STATISTICAL ENERGY ANALYSIS

Since its introduction in the early 1960s, statistical energy analysis, or SEA as it is commonly called, has gained acceptance as a method of analysis for structural–acoustic systems.[1,2] Statistical energy analysis draws on many of the fundamental concepts from statistical mechanics, room acoustics, wave propagation, and modal analysis.[3–8] At first, SEA appears to be a very simple method of analysis. However, because of the diversity of concepts used in formulating the basic SEA equations, the method quickly becomes very complex. For this reason, analysts have recommended caution in using SEA. However, when used properly, SEA is a powerful method of vibration and acoustic analysis.

In SEA, the system being analyzed is divided into a set of coupled subsystems. Each subsystem represents a group of modes with similar characteristics. The SEA subsystems can be considered to be "control volumes" for vibratory or acoustic energy flow. Under steady-state conditions, the time-average power input to a subsystem from external sources and from other connected subsystems must equal the sum of the power dissipated within the subsystem by damping and the power transmitted to the connected subsystems

Consider, for example, a piece of machinery located within an enclosure in a large equipment room, as shown in Fig. 4. The noise in the equipment room due to operation of the machine is of concern. A simple SEA model for this problem is shown in Fig. 5. In this model three subsystems are used: one for the acoustic modes of the interior space within the enclosure; one for bending modes of the enclosure walls; and one for the acoustic modes of the exterior space in the equipment room. The airborne and structure-borne noise from the machine are specified as power inputs to the model. The input power to the enclosure acoustic space, W_a^{in}, is taken to be the airborne noise radiated by the machine, which can be determined using acoustic intensity measurements. The input power to the enclosure wall, W_s^{in}, is determined from the vibration of the machine at its attachment to the enclosure base. The time-average power dissipated within each subsystem is indicated by the terms W_a^{diss}, W_s^{diss}, and W_r^{diss}.

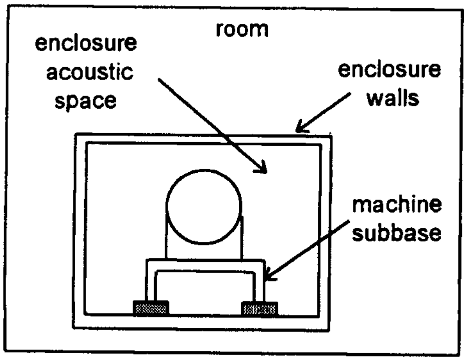


Fig. 4 Machinery noise problem.

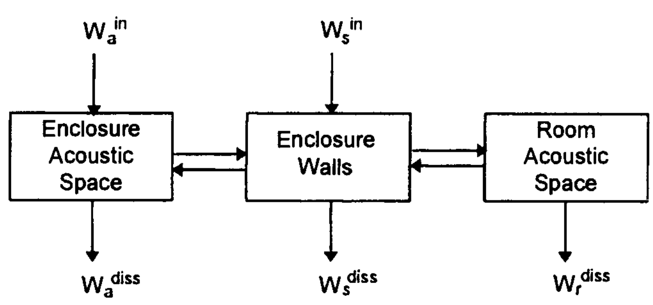


Fig. 5 Simple three-element SEA model of equipment enclosure.

Following the usual definition of the damping loss factor, the time-average power dissipated within the subsystem can be written in terms of the time-average energy of the system and the radian frequency of vibration,

$$W_s^{\text{diss}} = \omega \eta_{s;\,\text{diss}} E_s, \tag{18}$$

where ω is the radian frequency (typically, a one-third-octave band center frequency), $\eta_{s;\,\text{diss}}$ is the damping loss factor for subsystem s, and E_s is the time-average energy for subsystem s.

The energy transmitted between the connected subsystems can also be assumed to be proportional to the energy in each system. By analogy to the dissipated power, the factor of proportionality for transmitted power is called the coupling loss factor. However, since energy flow between the two systems can be in either direction, two coupling loss factors must be identified, so that the net energy flow between two connected subsystems is given by

$$W_{a;s}^{\text{trans}} = \omega \eta_{a;s} E_a - \omega \eta_{s;a} E_s, \tag{19}$$

where $\eta_{a;s}$ and $\eta_{s;a}$ are the coupling loss factors between subsystem a and s and between s and a. These two coupling loss factors are not equal. A power balance can now be performed on each subsystem to form a set of linear equations relating the energies of the subsystems to the power inputs.

$$\omega \begin{bmatrix} \eta_{a;d} + \eta_{a;s} + \eta_{a;r} & -\eta_{s;a} & -\eta_{r;a} \\ -\eta_{a;s} & \eta_{s;d} + \eta_{s;a} + \eta_{s;r} & -\eta_{r;s} \\ -\eta_{a;r} & -\eta_{s;r} & \eta_{r;d} + \eta_{r;a} + \eta_{r;s} \end{bmatrix} \cdot \begin{bmatrix} E_a \\ E_s \\ E_r \end{bmatrix} = \begin{bmatrix} W_a^{\text{in}} \\ W_s^{\text{in}} \\ W_r^{\text{in}} \end{bmatrix}. \tag{20}$$

Note that the subscript notation typically used in SEA is not conventional matrix notation. Also note that the loss factor matrix is not symmetric.

3.1 SEA Reciprocity

The coupling loss factors used in SEA are generally not reciprocal, that is, $\eta_{s;r} \neq \eta_{r;s}$. If it is assumed, however, that the energies of the modes in a given subsystem are equal, at least within the concept of an ensemble average, and that the responses of the different modes are uncorrelated, a reciprocity relationship can be developed. The assumptions for this relationship are more restrictive than required for a general statement of reciprocity, so that the term SEA reciprocity should be used.

Statistical energy analysis reciprocity requires that the coupling loss factors between two subsystems be related by the modal densities.

$$n(\omega)_s \eta_{s;r} = n(\omega)_r \eta_{r;s}. \qquad (21)$$

Using this relationship, a new coupling factor, β, can be introduced that allows the energy balance equations to be written in a symmetric form,

$$\omega \begin{bmatrix} \beta_{a;d} + \beta_{a;s} + \beta_{a;r} & -\beta_{a;s} & -\beta_{a;r} \\ -\beta_{a;s} & \beta_{s;d} + \beta_{s;a} + \beta_{s;r} & -\beta_{s;r} \\ -\beta_{a;r} & -\beta_{s;r} & \beta_{r;d} + \beta_{r;a} + \beta_{r;s} \end{bmatrix} \cdot \begin{bmatrix} \dfrac{E_a}{n(\omega)_a} \\ \dfrac{E_s}{n(\omega)_s} \\ \dfrac{E_r}{n(\omega)_r} \end{bmatrix} = \begin{bmatrix} W_a^{\text{in}} \\ W_s^{\text{in}} \\ W_r^{\text{in}} \end{bmatrix}. \qquad (22)$$

where

$$\beta_{s;r} = \omega \eta_{s;r} n(\omega)_s = \omega \eta_{r;s} n(\omega)_r = \beta_{r;s}. \qquad (23)$$

The ratio of total energy to modal density has the units of power and can be called "modal power."

3.2 Coupling Loss Factor Measurement

The coupling loss factor, or coupling factors, cannot be measured directly. However, a power injection technique can be used to infer the coupling factor from measured values of power input and response energy. Using this technique, each subsystem is excited in turn with a unit power input, and the response energy of the subsystems is measured to form a matrix of measured energies. Each column in the matrix corresponds to the measured response energies when one subsystem is excited. For example, the second column contains the measured energies when the second subsystem is excited. The coupling loss factor matrix is determined by inverting the matrix of measured subsystem energies,

$$[\eta] = [E]^{-1}. \qquad (24)$$

The off-diagonal terms are the negative values of the coupling loss factors, while the sum of terms for each row gives the damping loss factors. This measurement technique has been successfully used to "measure" in situ coupling and damping loss factors. However, errors in the energy measurement can result in large errors in the measured loss factors. Systems containing highly coupled subsystems will result in energy matrices that are poorly conditioned, since two or more columns will be nearly equal. Thus, the success of the measurement technique requires careful identification of the subsystems. The best results are obtained for light coupling, when the coupling loss factors are small compared to the damping loss factors, so that the loss factor matrix is diagonally dominant.

The measurement of subsystem energy is particularly difficult for subsystems with in-plane compression and shear modes. Because of the high stiffness of the in-plane modes, a small amount of motion results in a large amount of energy. The measurement of subsystem energy is also difficult for subsystems in which the mass is nonuniformly distributed. For these subsystems an effective or dynamic mass must be determined at each measurement point.

3.3 Coupling Loss Factor Theory

Coupling loss factors can be predicted analytically using wave and mode descriptions of the subsystem vibrations. Waves are used when the number of dimensions of the subsystem is greater than the number of dimensions of the connection: for example, a beam connected at a point, a plate connected at a point or along a line, and an acoustic space connected at a point, line, or area. Modes are used when the number of dimensions of the subsystem is equal to the number of dimensions of the connection: for example, a beam connected along a line and a plate connected over an area.

When a wave description can be used for all subsystems at the connection, the coupling loss factor between subsystems can be written in terms of a power transmission coefficient. For a point connection between beams, the coupling factor between subsystem s and subsystem r can be written as

$$\beta_{s;r} = \omega \eta_{s;r} n_s(\omega) = \frac{1}{2\pi} \tau_{s;r}, \qquad (25)$$

where $\tau_{s;r}$ is the power transmission coefficient. The power transmission coefficient must take into account energy transmitted by all degrees of freedom at the connection: three translational degrees of freedom and three rotational degrees of freedom. For a point connection with a single degree of freedom (all other degrees of freedom are constrained), the transmission coefficient is given by

$$\tau_{s;r} = \frac{4R_s R_r}{|Z_j|^2}, \tag{26}$$

where R is the subsystem resistance (real part of the impedance) for the unconstrained degree of freedom and Z_j is the junction impedance—the sum of the impedances of all subsystems connected at the point. The coupling factor given by Eqs. (25) and (26) can also be used for two- and three-dimensional subsystems connected at a point with a single degree of freedom, as long as the correct impedances are used. For point connections with multiple degrees of freedom, an estimate of the coupling factor can be obtained by summing the power transmission coefficients for each degree of freedom.

The coupling factor between two-dimensional subsystems connected along a line of length L can also be written in terms of a power transmission coefficient. However, for this case an integration must be performed over all angles of incidence. The coupling factor is given in terms of the angle-averaged transmission coefficient as

$$\beta_{s;r} = \omega\eta_{s;r} n_s(\omega) = \frac{1}{2\pi}\,\frac{k_s L}{\pi}\,\overline{\tau_{s;r}} \tag{27}$$

where k_s is the wavenumber of the source subsystem and $\overline{\tau_{s;r}}$ is given by

$$\overline{\tau_{s;r}} = \frac{1}{2}\int_{-\pi/2}^{+\pi/2} \tau_{s;r}(\theta_s)\cos(\theta_s)\, d\theta_s \tag{28}$$

and θ_s is the angle of incidence for a wave in the source subsystem. The parameter $k_s L/\pi$ is the effective number of points for the line connection. For a line connection, the power transmission coefficent must take into account the energy transmitted by four degrees of freedom: three translational and one rotational. For a single degree of freedom, the transmission coefficient for an incident angle θ_s can be expressed in terms of the line impedances of the source and receiver subsystems as

$$\tau_{s;r}(\theta_s) = \frac{4R_s(k_r)R_r(k_t)}{|Z_j(k_t)|^2}, \tag{29}$$

where k_t is the trace wavenumber given by $k_s\cos(\theta_s)$ and $R(k_t)$ is the real part of the line impedance for the unconstrained degree of freedom.

The formulation above can also be used to predict the coupling loss factor between three-dimensional subsystems coupled along a line if the integration is performed over all solid angles of incidence. For this case, the angle-averaged transmission coefficient is written as

$$\overline{\tau_{s;r}} = \int_{-\pi/2}^{+\pi/2} \tau_{s;r}(\theta_s)\sin(\theta_s)\cos(\theta_s)\, d\theta_s. \tag{30}$$

For an area connection between three-dimensional subsystems, the coupling factor is given in terms of the angle-averaged transmission coefficient as

$$\beta_{s;r} = \omega\eta_{s;r} n_s(\omega) = \frac{1}{2\pi}\,\frac{k_s^2 S}{4\pi}\,\overline{\tau_{s;r}}, \tag{31}$$

where S is the area of the connection. The effective number of points for the area connection is given by the parameter $k_s^2 S/4\pi$.

When the number of dimensions of a subsystem is equal to the number of dimensions of the coupling, modes are used to calculate the coupling loss factor. For example, the coupling loss factor between a two-dimensional system such as a plate or shell and a three-dimensional system such as an acoustic space is obtained by calculating the radiation efficiency for each mode of the plate and averaging over all modes with resonance frequencies in the analysis bandwidth,

$$\eta_{s;r} = \frac{\rho_r c_r S}{\omega M_s}\,\frac{1}{N_s}\sum_i \sigma_i^{\text{rad}}, \tag{32}$$

where $\rho_r c_r$ is the characteristic impedance of the acoustic space, M_s is the mass of the plate, N_s is the mode count for the plate, and σ_i^{rad} is the radiation efficiency for mode i of the plate. Approximations to the summation can be made by grouping the modes into "edge" and "corner" modes.[9]

The power transmission coefficients and radiation efficiencies can be calculated with great accuracy. However, the relationship between these parameters and the SEA coupling loss factors requires that some assumptions be made regarding the vibration fields in the connected subsystems. First, the vibrations of the two subsystems are assumed to be uncorrelated. Second, the vibrations of the two subsystems are assumed to be "diffuse"—waves are incident on a point within the subsystem from all angles with equal intensity. Although these assumptions are difficult to prove, even for idealized structures and acoustic spaces, they are generally valid for lightly coupled systems at high frequencies, where many modes participate in the vibration response. The validity of the assumptions for highly coupled subsystems is open to question. Fortunately, the errors incurred using the above equations for highly coupled subsystems are generally small. The assumptions are also open to question at low frequencies, where only a few modes participate in the response. At these frequencies, the equations above may predict coupling loss factors that are too large. However, it is difficult to quantify the error. In spite of the limited validity of the assumptions, the equations above provide useful estimates of the coupling loss factor, even for highly coupled subsystems at low frequencies.

3.4 Damping Loss Factor Theory

The damping loss factors can be predicted analytically for free-layer and constrained-layer treatments. The analysis approach is described in Chapter 71. The damping for an acoustic space is often specified by the average absorption coefficient for the space rather than a damping loss factor. The power dissipated within the acoustic space can be written in terms of the time-average energy and the absorption coefficient as

$$W_a^{\mathrm{diss}} = \omega \frac{S_a}{4k_a V_a} \alpha_{a;\mathrm{diss}} E_a = \omega \eta_{a;\mathrm{diss}} E_a, \tag{33}$$

where S_a is the area of the absorbing surface, V_a is the volume of the acoustic space, and k_a is the acoustic wavenumber. It follows that the damping loss factor for an acoustic space can be obtained from the average absorption coefficient by the relation

$$\eta_{a;\mathrm{diss}} = \frac{S_a}{4k_a V_a} \alpha_{a;\mathrm{diss}}, \tag{34}$$

where the constant $4V/S$ is commonly referred to as the mean free path.

3.5 Energy and Response

The SEA power balance equations can be solved to obtain the modal energy or modal power of each subsystem. The final step in the analysis is to relate these variables to the subsystem response. For structural subsystems, the spatial-average mean-square velocity is calculated from the kinetic energy. For resonant vibrations, the time-average kinetic energy is equal to the time-average potential energy. Thus, the average mean-square velocity in a band of frequencies is given by

$$\langle v^2 \rangle_{x,t} = \frac{E}{M}, \tag{35}$$

where E is the total energy of all modes in the band and M is the mass of the subsystem. For acoustic subsystems, the spatial-average mean-square acoustic pressure is calculated from the potential energy,

$$\langle p^2 \rangle_{x,t} = \frac{E}{C_a}, \tag{36}$$

where C_a is the compliance of the subsystem, $V/\rho c^2$, for an acoustic space with rigid walls.

3.6 Variance

Statistical energy analysis provides a statistical description of the subsystem response. However, in many cases, SEA is used only to obtain an estimate of the mean. Although the mean provides the "best estimate" of the response, this estimate may differ significantly from the response measured for a single member of the ensemble of dynamic systems. The variance or standard deviation of the response provides a method to quantify the expected confidence intervals for the SEA prediction. When the variance is high, the confidence intervals will be large, and the mean does not provide an accurate estimate of the response.

Using SEA in design requires that a confidence interval be established for the response prediction, so that an upper bound or "worst-case" estimate can be compared with design requirements. If the mean response is used for design, half the products produced will fail to meet the design requirements. Use of the mean plus 2 times the standard deviation (square root of the variance) provides a reasonable upper bound for the response prediction.

Methods to predict the variance of the SEA prediction are not well established. Often an empirical estimate of the variance or confidence interval is used. In other cases, an estimate based on the modal overlap parameter and the frequency bandwidth of the analysis is used. The modal overlap parameter M_{overlap} is the ratio of the average damping bandwidth for an individual mode to the average spacing between resonance frequencies. This parameter can be written in terms of the damping loss factor and the modal density as

$$M_{\mathrm{overlap}} = \tfrac{1}{2}\pi\omega\eta_d n(\omega),$$

where η_d is the effective total damping loss factor for the subsystem. Large values of the product of the modal overlap parameter and the analysis bandwidth result in low variance and a narrow confidence interval. In this case, the mean is a good estimate of the response. Small values of the product result in high variance and wide confidence intervals. In this case, the mean does not give a good estimate of the response, and the variance should be determined so that an upper bound to the prediction can be obtained.

Failure to include an estimate of the variance in the SEA analysis leads to some misunderstandings regarding the capabilities of SEA. First, SEA is not limited to high frequencies and high modal densities. However, at low frequencies and for low modal densities, the confidence interval for the SEA predictions will be large, so that an estimate of the variance must be made. Second, SEA is not limited to broadband noise analysis. However, for a single-frequency or narrow-band analysis, the confidence interval for the SEA predictions will be larger than for a one-third-octave or octave band analysis.

REFERENCES

1. J. E. Manning, "Statistical Energy Analysis—An Overview of Its Development and Engineering Applications," *Proc. 59th Shock and Vibration Symposium*, Albuquerque, New Mexico, 1988, pp. 25–38.
2. K. H. Hsu and D. J. Nefske (Eds.), *Statistical Energy Analysis*, Vol. 3, Am. Soc. of Mech. Eng., NCA, New York, 1987.
3. R. H. Lyon and R. G. DeJong, *Theory and Application of Statistical Energy Analysis*, 2nd ed., Butterworth-Heinemann, Newton, MA, 1995.
4. M. P. Norton, "Statistical Energy Analysis of Noise and Vibration," in *Fundamentals of Noise and Vibration Analysis for Engineers*, Cambridge University Press, Cambridge, U.K., 1989, Chap. 6.
5. F. J. Fahy, "Statistical Energy Analysis," in R. G. White and J. G. Walker (Eds.), *Noise and Vibration*, Ellis Horwood, Chichester, U.K., 1982, Chap. 7.
6. I. L. Ver, "Statistical Energy Analysis," in L. L. Beranek and I. L. Ver (Eds.), *Noise and Vibration Control Engineering Principles and Applications*, Wiley, New York, 1992, Sect. 9.8.
7. E. E. Ungar, "Statistical Energy Analysis," in L. Cremer, M. Heckl, and E. E. Ungar (Eds.), *Structure-Borne Sound*, 2nd ed., Springer-Verlag, Berlin, 1988, Sect. V.8.
8. W. G. Price and A. J. Keane (Eds.), "Statistical Energy Analysis," *Phil. Trans. Royal Soc. London*, Vol. 346, No. 1681, 1994, pp. 429–552.
9. G. Maidanik, "Response of Ribbed Panels to Reverberant Acoustic Fields," *J. Acoust. Soc. Am.*, Vol. 34, 1962, pp. 809–826.

PART VIII

NOISE: ITS EFFECTS AND CONTROL

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker

Copyright © 1997 John Wiley & Sons, Inc.

79

NOISE CONTROL

MALCOLM J. CROCKER

1 INTRODUCTION

Noise is usually defined as unwanted sound. Its main effects on people include interference with speech and sleep. Noise causes annoyance and if sufficiently intense it can cause hearing loss (Chapter 119). In the United States it is estimated that 5.5 million workers in manufacturing and utilities are exposed to noise conditions that are hazardous to hearing.[1] According to a survey in Germany in 1994, 70% of the citizens are annoyed by environmental noise caused by road traffic and a further 40% by aircraft. In addition over 20% are annoyed by rail traffic and industrial noise.[2,3] Noise is normally considered in the *source–path–receiver* framework.[4] Part VIII describes some of the many machinery noise sources (Chapters 83 and 86) and methods of control, both passive (Chapter 83) and active (Chapter 84). The ways in which noise effects people and criteria for determining human response to noise are discussed in Chapter 80. Methods of measuring noise are discussed briefly in this part, both in Chapters 82 and 86. In most developed countries the major noise problems are industrial noise, aircraft and airport noise, surface transportation noise, and community noise. These topics have received special coverage in Chapters 86–88. In addition the control of noise and vibration sources in buildings are described in Chapter 85. In cases where noise is very intense (and cannot be reduced sufficiently by engineering controls) the use of hearing protection devices must be considered (see Chapter 81). The control of noise requires a good understanding of many concepts including sound sources, both steady state and transient (Chapter 9 and 10) and sound propagation inside buildings (Chapter 92) and through the atmosphere (Chapter 32). The reader, concerned with noise, will also find many other chapters in this book useful including those concerned with the ear and hearing (Parts XI and XII) and those in Part XVI, where acoustical measurements (Chapter 147), sound intensity measurements (Chapter 150) and analyzers (Chapter 151) are described. Since vibration often results in airborne noise, readers will also find the discussions in Part VI useful, including vibration measurements and instrumentation (Chapter 72) and vibration isolation and damping (Chapter 71). For readers who require further in-depth coverage of noise problems and methods for their solution, there are several books that give a comprehensive treatment.[4–8]

2 CRITERIA AND RATING MEASURES FOR NOISE

Noise disturbs people in a variety of ways. If it is sufficiently intense, it can cause hearing damage. Very intense noise can fracture the eardrum and cause immediate damage. If the noise is less intense but experienced over a considerable period of time, such as in some industrial situations, it will also cause damage.[1] This situation is described in Chapter 119. Noise has other effects too. It interferes with speech and communications in general, spoils sleep, and causes annoyance.[9] A large number of noise measures have been produced to evaluate noise. The effects of noise are to a certain extent dependent on level, duration, frequency content, variation in level with time, presence of pure tones in the noise, existence of background noise, and so on. Thus some of the measures make allowance for some of these factors. There are a bewildering number of measures that have been proposed to evaluate noise. Some are no longer in

---

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker
Copyright © 1997 John Wiley & Sons, Inc.

use. The search to find one single measure that could be used to evaluate a large number of noise situations has largely been abandoned. However the day–night equivalent sound level, usually A-weighted, has become very widely used to evaluate community noise particularly near airports. Chapter 80 describes a number of these noise measures that have been widely used. Included are some measures that have gone out of favor, but from which other measures have developed and for which a large amount of data exist in the literature.

3 HEARING PROTECTION DEVICES

As already discussed, hearing damage can be experienced in situations when the noise is intense. Hearing protectors offer a "first line of defense," but should really only be used as a last resort when engineering and other control measures fail. Hearing protectors come in three main types: earplugs, earmuffs, and helmets. Although they have been used for thousands of years (Odysseus used wax earplugs to prevent him from hearing the Sirens' song), they are now mainly used to reduce industrial noise damage or to cut out noise to aid in sleeping. Earmuffs usually consist of rigid earcups that contain acoustic absorbing material and are held in place against the head by a flexible band. The earmuffs avoid the unpleasant pressure on the ear canal made by earplugs, but can be uncomfortable to wear in hot weather. Helmets are sometimes used. Helmets not only protect the head against impacts, but also provide ear protection as well since they frequently contain internal earcups. All hearing protection devices should be fitted carefully to achieve full hearing protection. If leaks exist in earplugs, the degradation in performance can be considerable. Likewise leaks with earmuffs, caused for instance by hair or glasses that break the earmuff seal, also reduce performance. Chapter 81 describes the main types of hearing protectors and also their fitting and attenuation. This chapter also describes some of the many ways available to measure the attenuation provided by hearing protectors and provides information on the effects of hearing protectors on speech intelligibility, the ability to hear warning signals, and localization of sound.

4 NOISE EMISSION OF SOUND SOURCES

To determine the acceptability of the noise produced by a noise source, its noise emission must be determined. The noise emitted by moving sources such as aircraft or vehicles is usually described by the sound pressure level at a certain defined location. Stationary sources, however, particularly those which operate in a well-defined environment or are small enough to be moved into one, are usually described by their sound power output. Chapter 82 describes methods to determine the sound power of a source from sound pressure measurements. Several are available. If the source is small enough and portable, it can be placed in a reverberation room if one is available and then an accurate estimate of its sound power can be obtained. Alternatively, if the source cannot be moved and only its normal environment rather than a reverberant room is available, then its sound power can still be estimated from sound pressure measurements made on a hemispherical or rectangular enveloping measurement surface. Since the sound pressure measured on the enveloping surface will be increased by background noise and/or reverberation, a correction factor must be applied in such cases. The correction factor to account for reverberation can be determined by measuring the reverberation time. The use of a reference sound source, however, can enable the correction factor for either background noise or reverberation (or both) to be obtained by using a substitution method. If the background noise is too high, then the correction is inaccurate and can no longer be applied. In such cases it is advisable to obtain the sound power of the source using sound intensity measurements (see Chapter 156). Accurate sound power determinations can be made in cases where the background noise is much higher than is possible using the mean square sound pressure approach described in Chapter 82. Of course the sound intensity measurement equipment may not be available, or it may be considered to be too expensive to use.

5 NOISE GENERATION IN MACHINERY AND ITS CONTROL

If the sound pressure level or the sound power level of a source has been determined and found to exceed certain criteria, then it may be decided to reduce the noise emission of the source by engineering methods. At the end of Chapter 83, methods to identify sources of noise in a complicated machine or system are described. These methods include selective operation, selective lead wrapping, frequency analysis, sound intensity measurements, multiple coherence measurements, and cepstrum among others. It is normally necessary to use more than one method to obtain more confidence in the results. After good information is obtained about the noise sources on a machine, then noise control methods can be used. Chapter 83 also describes the main passive noise control approaches that have been useful in reducing the noise of machinery. These include vibration isolation, use of sound absorbing materials, machinery enclosures, and use of barriers and application of damping materials. Although there are many different types of machines in existence, a few of the main types of machines or machine elements and their noise generating mechanisms are described in Chapter 83. Some of these machine noise discussions may be helpful even if the reader is faced with a noise problem with a different type of machine.

6 ACTIVE NOISE CONTROL

The main passive noise control approaches described in Chapter 83 (vibration isolation, absorption, enclosure, barriers, and damping), although well understood and usually implemented with only moderate cost, generally do not work very well at low frequency. This has led to the search for alternative methods that can be used in the low-frequency region. Active noise control seems to offer a possible solution. In recent years low-cost digital signal processors have become available. Together with the wave nature of sound they have made active noise control feasible. In this approach the sound field is sensed with a microphone, processed by a microprocessor, and fed back through a loudspeaker so as to destructively interfere with the sound field of the primary source. An error microphone situated away from the source whose signal is minimized as much as possible can also be used. There are two main approaches that are described in this chapter: feedback control and feed forward control. In general it is easier to control pure-tone sources than broadband noise sources. Active control technologies are only now beginning to emerge from the laboratory and become used in real engineering applications. The design of active control systems still remains a task for specialists, however. Active noise control has been successfully applied to several practical cases such as (1) the low-frequency *broadband* noise in ear defenders, air conditioning ducts, the exhaust pipes of vehicle engines, and the interior noise inside road vehicles, and (2) the low-frequency *pure-tone* noise in the cabins of propeller-driven aircraft, engine-generated pure-tone noise in vehicle interiors, and the pure-tone noise generated by electrical transformers. Active vibration control has also begun to be applied to the control of vibration in the form of active vibration isolation and distributed sensors and actuators in extended structures such as plates and cylinders. PVDF layers have frequently been used as sensors and PZT as actuators. Of course structural systems are frequently coupled to acoustic systems, and if the structural vibration can be controlled so can the noise in the coupled acoustic system. Chapter 84 reviews both the theory for active noise control systems and practical applications.

7 NOISE CONTROL IN AIR CONDITIONING SYSTEMS

One of the main noise sources in buildings is the noise produced by the air conditioning system. The main source is of course the fan, but additional sound may also be generated by fittings and take-offs. Sound is naturally attenuated as it travels along ducts by transmission through duct walls (break-out), by energy division at take-offs, and by end reflection when the duct terminates at the room. If the sound transmitted to a room is too high, however, it can be reduced by lining the interior surfaces of the duct with sound-absorbing material or by using packaged attenuators or both. Care should be taken in air conditioning systems to minimize aerodynamic noise generation at elbows, turns, take-offs, silencers, and so on. The noise generated depends not only on the duct geometry, but also on the flow velocity. If the flow velocity is minimized, so also is the noise. Chapter 85 deals with the problems of air conditioning noise and, in particular, fan noise, duct attenuation, the effects of duct lining, turbulent flow generation, and sound generation and control in outlet devices. Chapter 98 also deals with air conditioning noise and, in particular, deals with the sound generated by the fan or blower.

8 SOUND POWER LEVELS OF MACHINERY

As discussed in Section 4, the sound power level is the preferred way to describe the noise emission of sources. Fixed machinery such as that used in industry is normally described by its sound power level. From this the sound pressure level in the interior of a building or at a certain specified distance from the source(s) in the community can be calculated. Chapter 86 presents procedures to calculate the sound power levels of equipment, including boilers, electric motors, gas turbines, reciprocating engines, steam turbines, transformers, compressors, pumps, fans, gears, generators, valves and piping, air conditioning equipment, cooling towers, chillers, construction equipment, and oil field equipment. The equations presented in Chapter 86 are either based on measured data or are semi-empirical. In construction projects, it is frequently necessary to make acoustical predictions. In that case either measurements should be made on the machinery to be used, or sound power information should be obtained from the manufacturers of the machinery. If neither of these options is available, then the procedures for estimating the sound power presented in this chapter can be used.

9 AIRCRAFT AND AIRPORT NOISE

Civilian air transport continues to expand worldwide. Although the latest passenger jets with their bypass turbofan engines are significantly quieter than the first generation of jet aircraft, which used pure turbojet engines, passenger aircraft jet noise remains a problem, particularly near airports. Although the noisiest passenger aircraft are being retired in some countries, airport noise is likely to remain a serious problem because in many countries the frequency of aircraft operations continues to increase and because of the public opposition to noise voiced by some citizens groups near to airports. This opposition has prevented extensions to some airports and the development of some other airports entirely because of environmental concerns. Chapter 80 describes some of the aircraft noise descriptors that have been used in the United States and overseas, such as CNR, NEF, and NNI. The NEF descriptor has now been replaced in the United States by the day–night equivalent sound level L_{dn}. However, a considerable variety of older varied descriptors remain in use in many other countries.[3] Chapter 87 discusses the way in which aircraft noise is monitored at airports in the United States and describes how to calculate airport noise and its impact on the population. Also discussed is noise abatement at the national level. Specific airport noise control measures that can be taken are described, including optimization of flight tracks, definition of preferential runways to minimize noise, setting of standards to restrict the noisiest aircraft, use of noise abatement procedures during take-off and landing to minimize noise. In addition, airport layout changes to reduce noise, soundproofing of schools, churches, and apartment buildings, and the erection of barriers to reduce the noise impact on the population are discussed.

10 SURFACE TRANSPORTATION NOISE

Surface transportation is really a bigger noise problem than aircraft noise. It affects more people than aircraft noise in most countries. This chapter describes noise sources on a diesel engine truck. Chapter 88 also discusses vehicle noise sources and describes in considerable detail noise regulations and limits for vehicle noise that apply to new highway vehicles sold in the United States, Canada, the European Union, Japan, and other countries. Chapter 88 also describes test procedures to measure vehicle noise. These test procedures normally require the vehicle to be operated in a full-acceleration condition, but in Japan fixed-speed passby tests at 60% rated engine speed and 65 km/hr and a stationary test to determine exhaust noise are also required. The chapter also includes discussion on locomotive noise regulations and on interior vehicle noise.

11 COMMUNITY NOISE

Noise remains a serious problem in communities. Chapter 80 discusses some of the descriptors that have been used. There has been a move to the use of the A-weighted L_{eq} in many countries for the assessment of road traffic noise, although L_{10} remains in use in Australia, Hong Kong, and the United Kingdom for target values and insulation regulations for new roads and planning values for new residential areas.[3] L_{dn} has become the main descriptor of community noise in the United States since its adoption by several federal agencies in the 1970s. Chapter 89 describes the current practice of evaluating community noise problems in the United States. Shaw has written a useful recent review of community noise.[10]

REFERENCES

1. A. H. Suter and D. L. Johnson, "Programs in Controlling Occupational Noise Exposure," *Noise Control Engineering Journal*, Vol. 44, 1996, pp. 121–126.
2. Institute fur proxisorientierte Sozial forschung, Meinungen zur Umweltpolitik 1994, Mannheim, Germany, 1994.
3. D. Gottlob, "Regulations for Community Noise," *Noise/News International*, Vol. 3, 1995, pp. 223–236.
4. C. M. Harris (Ed.), *Handbook of Acoustical Measurements and Noise Control*, 3rd ed., McGraw-Hill, New York, 1991.
5. L. L. Beranek and I. L. Ver (Eds.), *Noise and Vibration Control Engineering*, Wiley, New York, 1992.
6. L. H. Bell and D. A. Bell, *Industrial Noise Control*, 2nd ed., Marcel Dekker, New York, 1994.
7. C. H. Hansen and D. A. Bies, *Engineering Noise Control: Theory and Practice*, 2nd ed., E & FN Spon, 1996.
8. M. J. Crocker and A. J. Price, *Noise and Noise Control*, Vol. 1, CRC Press, Boca Raton, 1975, and M. J. Crocker and F. M. Kessler, *Noise and Noise Control*, Vol. II, CRC Press, Boca Raton, 1982.
9. H. E. von Gierke and K. McK. Eldred, "Effects of Noise on People," *Noise/News International*, Vol. 1, No. 2, 1993, pp. 67–89.
10. E. A. G. Shaw, "Noise Environments Outdoors and the Effects of Community Noise Exposure," *Noise Control Engineering Journal*, Vol. 44, No. 3, 1996, pp. 109–120.

80

RATING MEASURES, DESCRIPTORS, CRITERIA, AND PROCEDURES FOR DETERMINING HUMAN RESPONSE TO NOISE

MALCOLM J. CROCKER

1 INTRODUCTION

People are exposed to noise during daytime and nighttime hours. During the day the noise can interfere with various activities and cause annoyance, and at night it can affect sleep. Very intense noise can even lead to hearing damage (Chapter 119). In the daytime the activities most affected are communications that involve speech between individuals, speech in telephone communications, and speech and music on radio and television. If the noise is more intense, it is normally more annoying, although there are a number of other attitudinal and environmental factors too that affect annoyance. There are a large number of ways to measure and evaluate noise, each normally resulting in a different noise measure, descriptor, or scale. The different measures and descriptors mainly result from the different sources (aircraft, traffic, construction, industry, etc.) and the different researchers involved in producing them. From these measures and descriptors, criteria have been developed to decide on the acceptability of the noise levels for different activities. These criteria are useful in determining whether noise control efforts are warranted to improve speech communication, reduce annoyance, and lessen sleep interference. This chapter contains a review and discussion of some of the most important noise measures and descriptors. In the past 20–30 years these measures and descriptors have undergone some evolution and change as researchers have attempted to find descriptors that best relate to different human responses and are more easily measurable with improved instrumentation. For completeness this evolution is traced and some measures and descriptors are described that are no longer in use since knowledge of them is needed in the study of the results of various noise studies reported in the literature.

2 LOUDNESS AND ANNOYANCE

As the level of the noise is increased, it is accompanied by an apparent increase in loudness. Loudness may be considered to be the subjective evaluation of the intensity of a noise when this evaluation is divorced from all the attitudinal, environmental, and emotional factors that may affect the listener's assessment of the annoying properties of the noise. Generally, if a noise is louder it is judged to be more annoying and vice versa, although there are exceptions. Table 1 shows some of the acoustical and nonacoustical factors that can contribute to the annoyance caused by noise. Some of the factors shown in Table 1 are also important in considerations of the effects of noise on speech communication (see Sections 6 and 7) and on sleep (see Section 15.1). The annoyance caused by noise is discussed further in Section 15.2.

3 LOUDNESS AND LOUDNESS LEVEL

As discussed in Chapter 118, which contains an in-depth discussion of the loudness of sound, the human ear does

---

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker
Copyright © 1997 John Wiley & Sons, Inc.

TABLE 1 Some Acoustical and Nonacoustical Factors That Contribute to Annoyance Caused by Noise

Acoustical Factors	Sound pressure level Frequency spectrum Duration Pure-tone content Impulsive character Fluctuation in level
Nonacoustical Factors	Time of day Time of year Necessity for noise Community attitudes Past experience Economic dependence on source

not have a uniform sensitivity to sound as its frequency is varied. Figures 2 and 3 of Chapter 118 show equal-loudness-level contours. These contours connect together pure-tone sounds that appear equally loud to the average listener. Similar contours have been determined experimentally for bands of noise instead of pure tones. The unit of (linear) loudness is the *sone*. A sone is defined as the loudness of a pure tone with a sound pressure level of 40 dB at 1000 Hz. A sound that is twice as loud is said to have a loudness of 2 sones and so on. The *loudness level* of a 1000 Hz pure tone of 40 dB is defined as 40 *phons*. If the 1000 Hz pure tone is raised in level by 10 dB, it appears to be about twice as loud to the average listener. Thus a doubling of the loudness has been defined to be equivalent to an increase in loudness level of 10 phons. The relationship between loudness S and loudness level P (for both pure tones and bands of noise) is thus given by Eq. (1) and shown in Fig. 1.

$$S = 2^{(P-40)/10}. \qquad (1)$$

The preceding discussion has concerned the loudness of pure tones. Equal loudness contours for bands of noise have been determined experimentally, independently, by

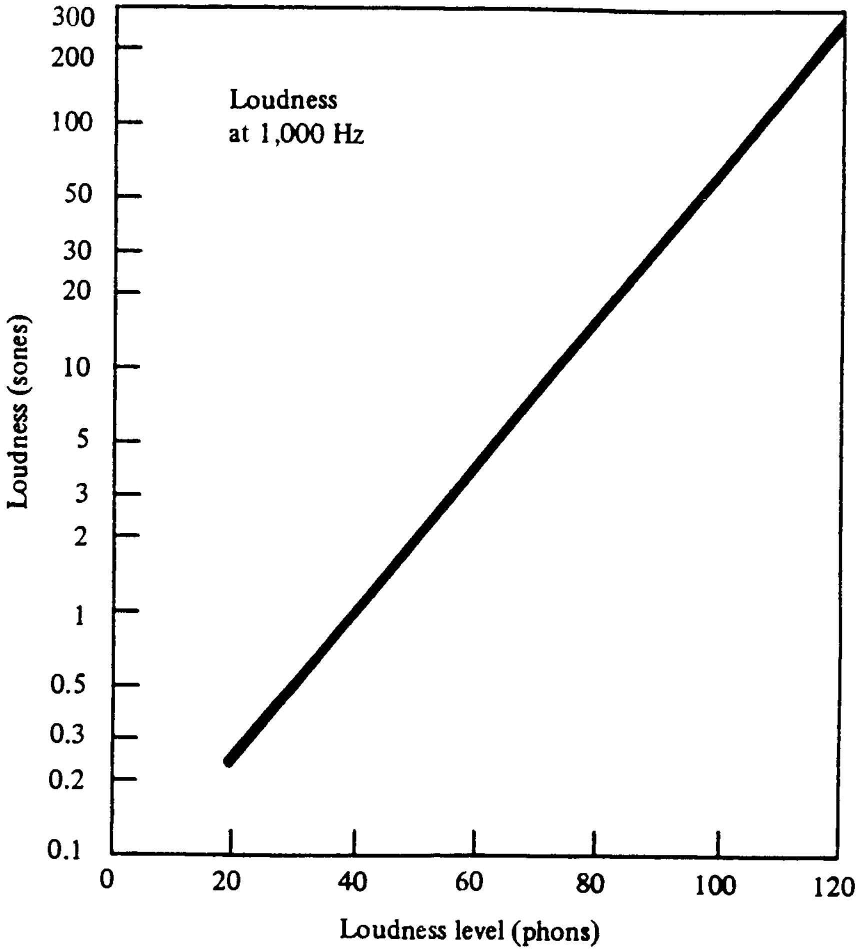


Fig. 1 The relationship between the loudness (in sones) and the loudness level (in phons) of a sound.

Stevens and Zwicker and standardized.[1–4] These can be used to evaluate the loudness (in sones) of noise sources. The procedure used in the Stevens mark VI method is to plot the noise spectrum in either octave or one-third-octave bands onto the loudness index contours. The loudness index (in sones) is determined for each octave (or one-third-octave) band and the total loudness S is then given by

$$S = S_{\max} + 0.3\left(\sum S - S_{\max}\right), \qquad (2)$$

where $S_{\max}$ is the maximum loudness index and $\sum S$ is the sum of all the loudness indices. The 0.3 constant (used for octave bands) in Eq. (2) is replaced by 0.15 for one-third-octave bands. The Stevens method assumes that the sound field is diffuse and does not contain any prominent pure tones. The Zwicker method is based on the critical band concept and, although more complicated than Stevens', can be used either with diffuse or frontal sound fields. Complete details of the procedures are given in the ISO standard[4] and Kryter[5] has discussed the critical band concept in his book.

4 NOISINESS AND PERCEIVED NOISE LEVEL

4.1 Noisiness

Although the level of noise or its loudness is very important in determining the annoyance caused by noise, there are other acoustical and nonacoustical factors that are also important. In laboratory studies, people were asked to rate sounds of equal duration in terms of their noisiness, annoyance, or unacceptability.[6,7] Using octave bands of noise, Kryter and others have produced *equal noisiness index* contours. These equal noisiness contours are similar to those for equal loudness, except that at high frequency less sound energy is needed to produce equal noisiness and at low frequency more is needed. The unit of noisiness index is the *noy* N. Equal noisiness index curves are shown in Fig. 2. The procedure to determine the logarithmic measure, the *perceived noise level* (PNL) is quite complicated and has been standardized.[8] It has also been described in several books.[9–11] Briefly, it may be stated as follows. Tabulate the octave band (or one-third-octave band) sound pressure levels of the noise. Calculate the noisiness index in noise for each band from Fig. 2. Then calculate the total noisiness index N_t from:

$$N_t = N_{\max} + 0.3\left(\sum N - N_{\max}\right) \qquad (3)$$

where $N_{\max}$ is the maximum noisiness index and $\sum N$ is the sum of all the noisiness indices. If one-third-octave bands are used, the constant 0.3 for octave bands in Eq. (3) is replaced by 0.15.

The total perceived noisiness index N_t (summed over all frequency bands) is converted to the perceived noise level PNL or L_{PN} from:

$$L_{\mathrm{PN}} = 40 + (33.22)\log N_t. \qquad (4)$$

The procedure is similar to that for calculating loudness level (phons) from loudness (sones). Some have questioned the usefulness of this procedure since listeners in laboratory experiments do not seem to be able to distinguish between (1) loudness or (2) noisiness and annoyance. However, this procedure has been widely used in assessing single-event aircraft noise. In the United States the Federal Aviation Administration (FAA) has adopted the effective perceived noise level for the certification of new aircraft.

4.2 Effective Perceived Noise Level

Noise which is of long duration is normally judged to be more annoying than noise of short duration. In addition, if the noise contains pure tones buried in a broadband noise spectrum, it is also judged to be noisier than without such tones. To account for these effects *the effective perceived noise level* (EPNL) or L_{EPN} has been defined as

$$L_{\mathrm{EPN}} = L_{\mathrm{PN}} + C + D \qquad (5)$$

where C is the correction factor for pure tones (between 0 and 3 dB) and D is a correction for duration.[9,10] The procedure for calculating L_{EPN} is quite complicated, and its description is beyond the scope of this chapter. It is fully described in standards and in some other books.[8,10,12]

5 SOUND LEVELS

In Chapter 118, Fig. 3 shows that the ear is most sensitive to sounds in the midfrequency range around 4000 Hz. It has a particularly poor response to sound at low frequency. It became apparent to scientists in the 1930s that electrical filters could be designed and constructed with a frequency response approximately equal to the inverse of these equal loudness curves. Thus A-, B-, and C-weighting filters were constructed to approximate the inverse of the 40-, 70-, and 90-phon contours (i.e. for low-level, moderate and intense sounds), respectively. (See Fig. 3.) In principle, then, these filters, if placed between the microphone and the meter display of an instrument such as a sound level meter should give some indication of the loudness of a sound (but for pure tones only).

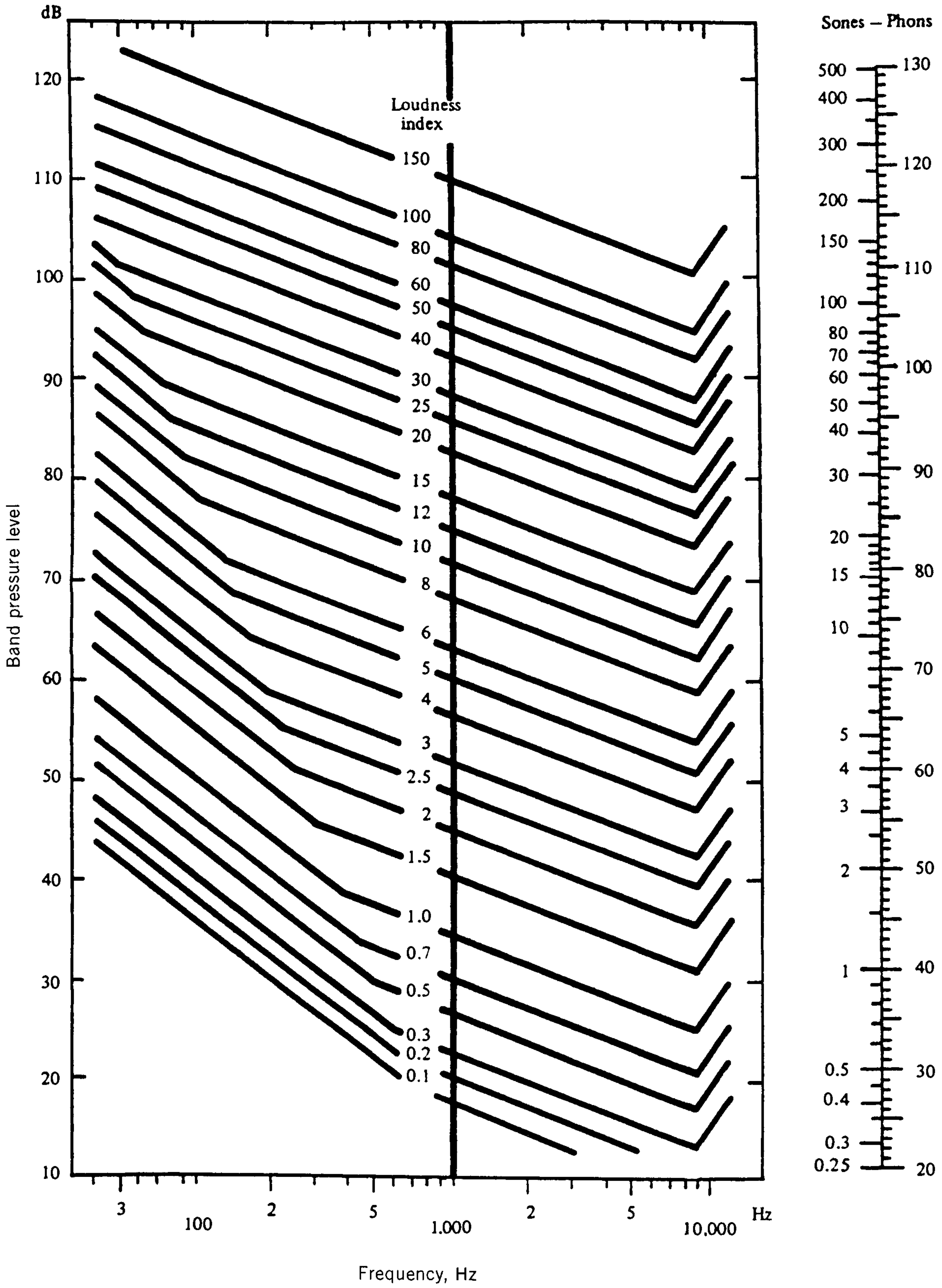


Fig. 2 Contours of equal loudness index.

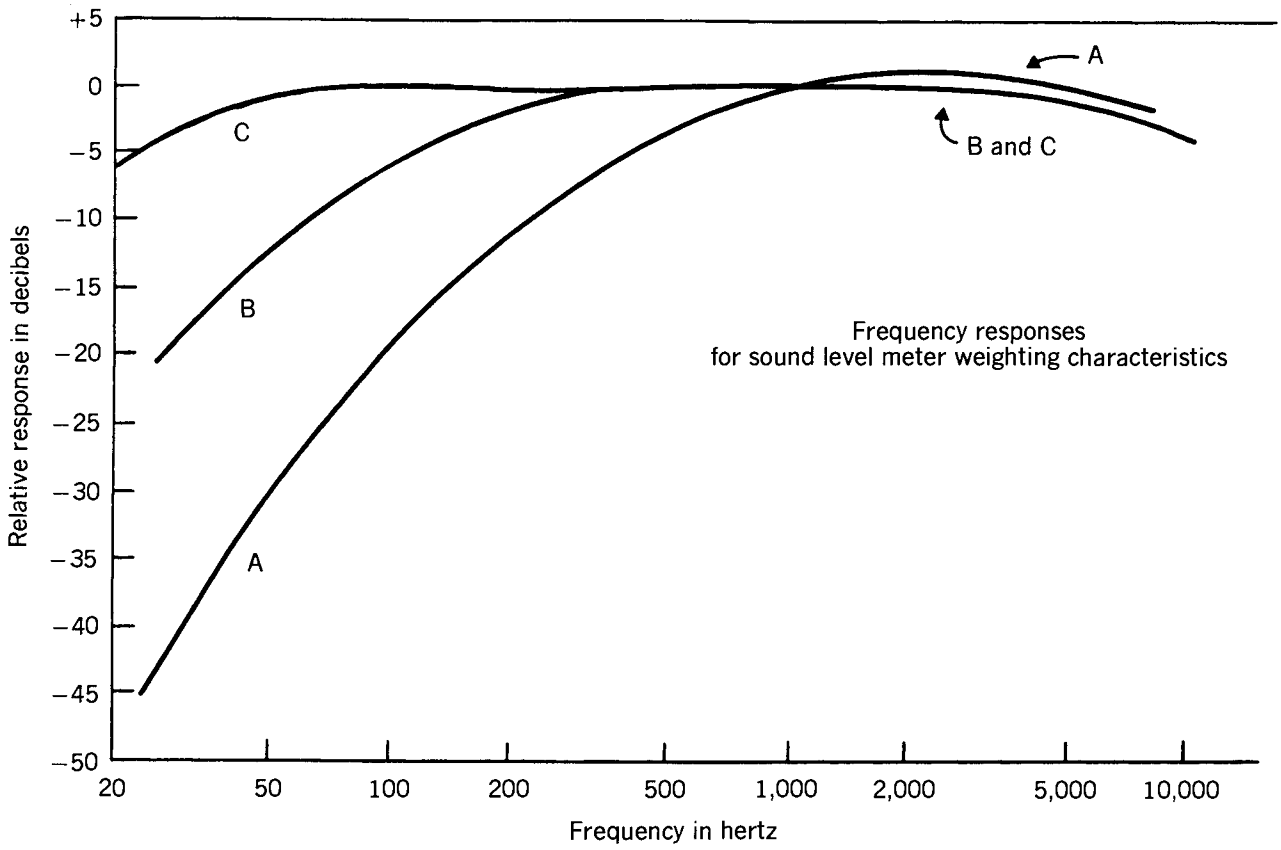


Fig. 3 A-, B-, and C-weighting filter characteristics used with sound level meters.

The levels measured with the use of these filters are commonly called the A-, B-, and C- weighted *sound levels*. The A-weighting filter has been much more widely used than either the B- or C-weighting filter and the *A-weighted sound pressure level* measured with it known simply as the *sound level* (unless the use of some other filter is specified). Because it is simple, giving a single number, and it can be measured with a low-cost sound level meter, the A-weighted sound level has been used widely to give an estimate of the loudness of noise sources such as vehicles, even though these produce moderate to intense noise. Beranek has reviewed the use of the A-weighted sound level as an approximate measure of loudness level.[13] The A-weighted sound levels are often used to gain some approximate measure of the loudness of broadband sounds and even of the acceptability of the noise. Figure 4 shows that there is reasonable correlation between the subjective response of people to vehicle noise and A-weighted sound levels for vehicle noise. The A-weighted sound level forms the basis of many other descriptors described later in this chapter and of the *sound exposure level* (SEL), *single-event noise exposure level* (SNEL), and *noise reduction* (NR) discussed in Chapter 94. The A-weighted sound level descriptor is also used as a limit for new vehicles (Chapter 88) and noise levels in buildings (Chapter 97) in several countries.

6 ARTICULATION INDEX

The *articulation index* (AI) is a measure of the intelligibility of speech in a continuous noise. The AI was first proposed by French and Steinberg[14] and was extended later by Beranek.[13] Speech has a dynamic range of about 30 dB in each one-third-octave band from 200 to 6000 Hz, and the long-term rms overall sound pressure level at the speaker's lips is about 65 dB. In speech, vowels and consonants are joined together to produce not only words but sounds that have a distinctive personal nature as well. The vowels usually have greater energy than consonants and give the speech its distinctive characteristics. This is due to the fact that vowels have definite frequency spectra with superimposed short duration peaks. The articulation index ranges from AI = 0 to 1.0 corresponding to 0% and 100% intelligibility, respectively. If the AI is less than about 0.3, speech communication is unsatisfactory (only about 30% of monosyllabic words understood);

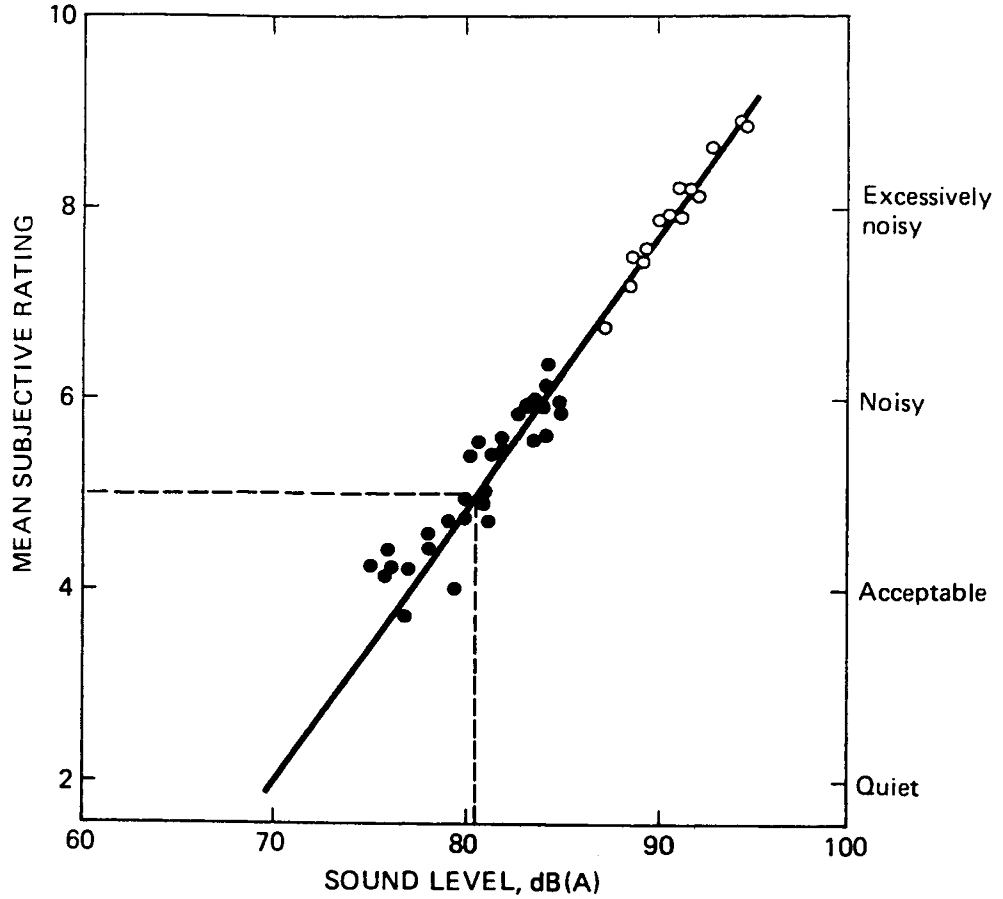


Fig. 4 Relation between subjective response and A-weighted sound level for diesel engine trucks undergoing an acceleration test; ● values measured in 1960, ○ values measured in 1968. (From C. H. G. Mills and D. W. Robinson, "The Subjective Rating of Motor Vehicle Noise," *The Engineer*, June 30, 1961; Ref. 39; and T. Priede, "Origins of Automotive Vehicle Noise," *J. Sound Vib.*, Vol. 15, No. 1, 1971, pp. 61–73.)

while if the AI is greater than about 0.6 or 0.7, speech communication is generally satisfactory (with more than 80% of monosyllabic words understood). Methods to calculate the AI are somewhat complicated and are given in an American National Standard[15] and explained in several books.[9,13] Since the calculation of AI is complicated, it will not be explained in detail here. The calculation of AI is described briefly in Chapter 94. The use of AI has somewhat declined in favor of the use of the *speech interference level.*

7 SPEECH INTERFERENCE LEVEL

The *speech interference level* (SIL) is a measure used to evaluate the effect of background noise on speech communication.[13] The speech interference level is the arithmetic average of the sound pressure levels of the interfering background noise in the four octave bands with center frequencies of 500, 1000, 2000, and 4000 Hz. (see ANSI S3.14-1977(R-1986)). If the SIL of the background noise is calculated, then this may be used in conjunction with Fig. 5 to predict the sort of speech required for satisfactory face to face communication with

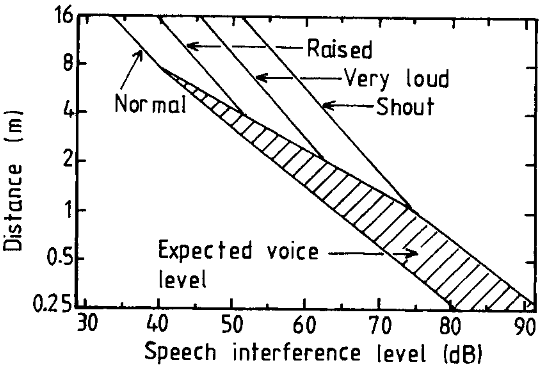


Fig. 5 Talker-to-listener distances for speech communication to be just reliable (reprinted from American Standard ANSI S3.14-1977).

male voices (i.e., for at least 95% sentence intelligibility). As an example, if the SIL is 40 dB and the speakers are males, they should be able to communicate with normal voices at 8 m. If the SIL increases to 50, raised voices must be used at 8 m. For females the SIL should be decreased by 5 dB (or the x-axis moved to the right by 5 dB.) The shaded area of Fig. 5 shows the range of speech levels that normally occur as people raise their voices to overcome the background noise. Speech interference level is also discussed in Chapter 94. Because it is simpler to measure than the SIL, the A-weighted sound level is sometimes used as a measure of speech interference, but with somewhat less confidence. Webster has produced a comprehensive diagram (see Fig. 6) which summarizes speech levels required for communication (at various distances) with 97% intelligibility of sentences for both outdoors and indoor situations.[16] Figure 6 is similar to Fig. 5, but contains some additional information concerning voice levels in different situations. With noise levels above 50 dB, people tend to raise their voice levels as shown by the "expected line" (at the left) for nonvital communication and the "communicating line" (at the right) of the diagonal shaded area for essential communication.

8 INDOOR NOISE CRITERIA

The speech interference level is mainly used to evaluate the effect of noise on speech in situations outdoors or indoors where the environment is not too reverberant. The A-weighted sound level can be used as a guide on the acceptability of noise in indoor situations, but it gives no indication which part of the frequency spectrum of the noise is of concern. A number of families of noise-weighting curves have been devised to evaluate the acceptability of noise in indoor situations. These include *noise criterion curves* (NC), *noise rating curves* (NR), *room criterion curves* (RC), and *balanced noise criterion curves* (NCB). The curves have resulted from the need to either specify acceptable noise levels in buildings or determine the acceptability of noise in existing build-

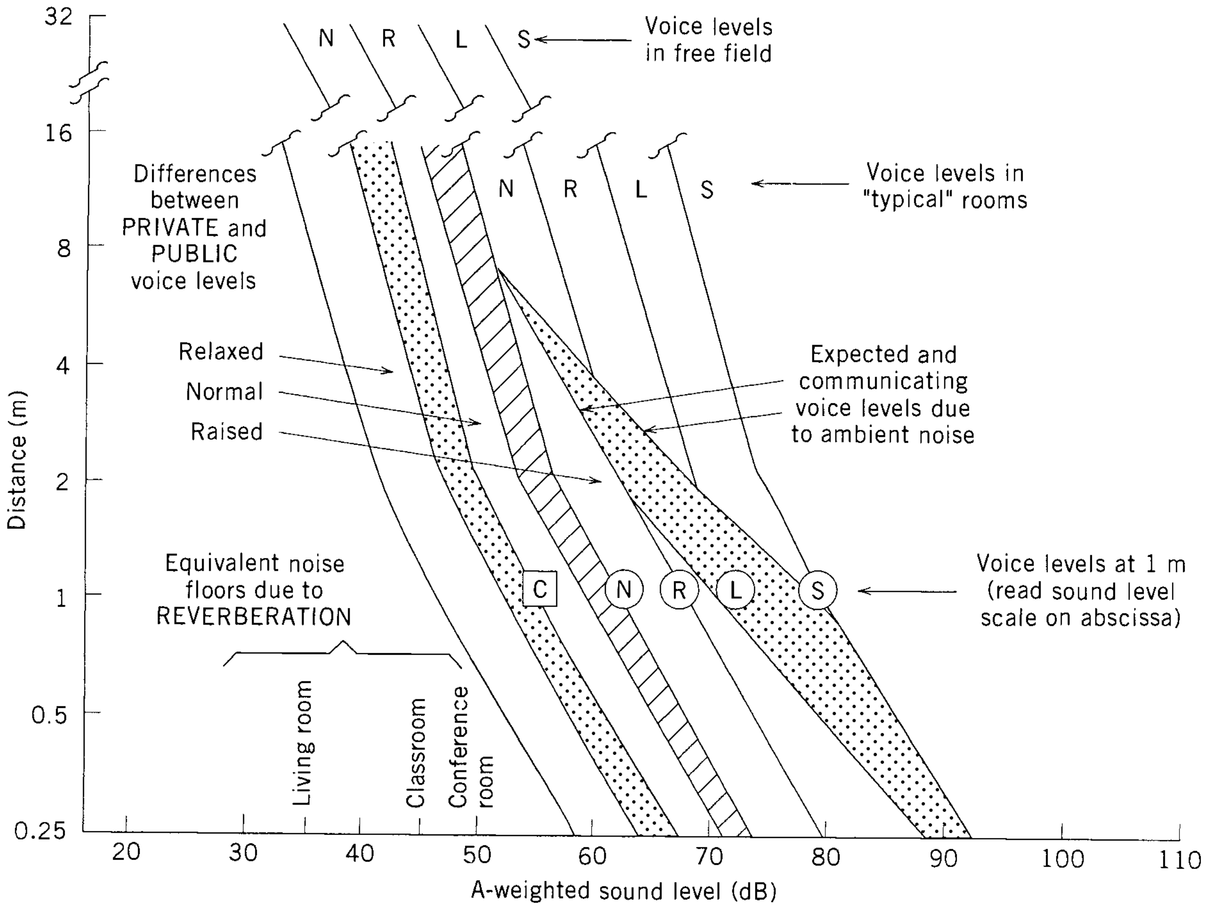


Fig. 6 Comprehensive diagram summarizing speech levels for communication (at various distances) for 97% intelligibility of sentences on first presentation to listeners for both outdoors (free field) and indoor situations.[16]

ing spaces. A major concern has been to determine the acceptability of air-conditioning noise. Beranek[17-20] has been a major contributor to the development of the NC and NCB noise criterion curves. Blazier[18,21] was mainly responsible for the development of the RC room criterion curves. NR curves were devised by Kosten and Van Os[22] and are similar to the NC curves. They have been standardized and adopted by the International Standards Organization. These noise-weighting curves are now reviewed briefly.

8.1 Noise Criterion Curves

The NC curves (Fig. 7) were developed from the results of a series of interviews with people in offices, public spaces, and industrial spaces.[17,18] These results showed that the main concern was the interference of noise with speech communication and listening to music, radio, and television. In order to determine the NC rating of the noise under consideration, the octave-band sound pressure levels of the noise are measured, and these are then plotted on the family of NC curves (Fig. 7). The noise spectrum must not exceed the particular NC curve specified in any octave band in order for it to be assigned that particular NC rating.[13] NC curves are also discussed in Chapter 94.

8.2 Noise Rating Curves

The *noise rating* NR curves are similar to the NC curves (see Fig. 7). They were originally produced to develop a procedure to determine whether noise from factories heard in adjacent apartments and houses is acceptable.[22] The noise spectrum is measured and plotted on the family of NR curves (Fig. 8) in just the same way as with the NC curves. One difference from the NC curves, however, is the use of corrections for time of day, intermittency, audible pure tones, fraction of time the noise is heard, and type of neighborhood. It has been found that in the range of NR or NC of 20 to 50 there is little difference between the results obtained from the two approaches.

8.3 Room Criterion Curves

NC curves are not defined in the low-frequency range (16- and 31.5-Hz octave bands) and are also generally regarded as allowing too much noise in the high-frequency region (at and above 2000 Hz). Blazier based his derivation of the *room criterion* RC curves on an extensive study conducted for ASHRAE by Goodfriend of generally acceptable background spectra in 68 unoccupied offices.[21] The A-weighted sound levels were mostly in the range of 40 to 50 dB. Blazier found that the curve

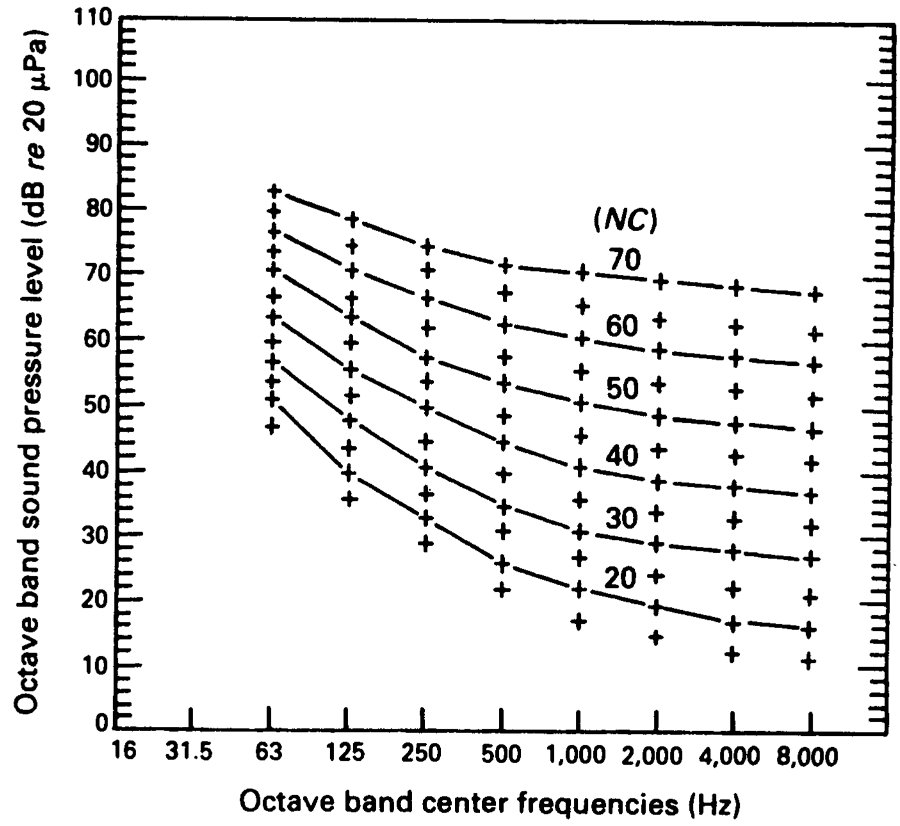


Fig. 7 Noise criterion (NC) curves.

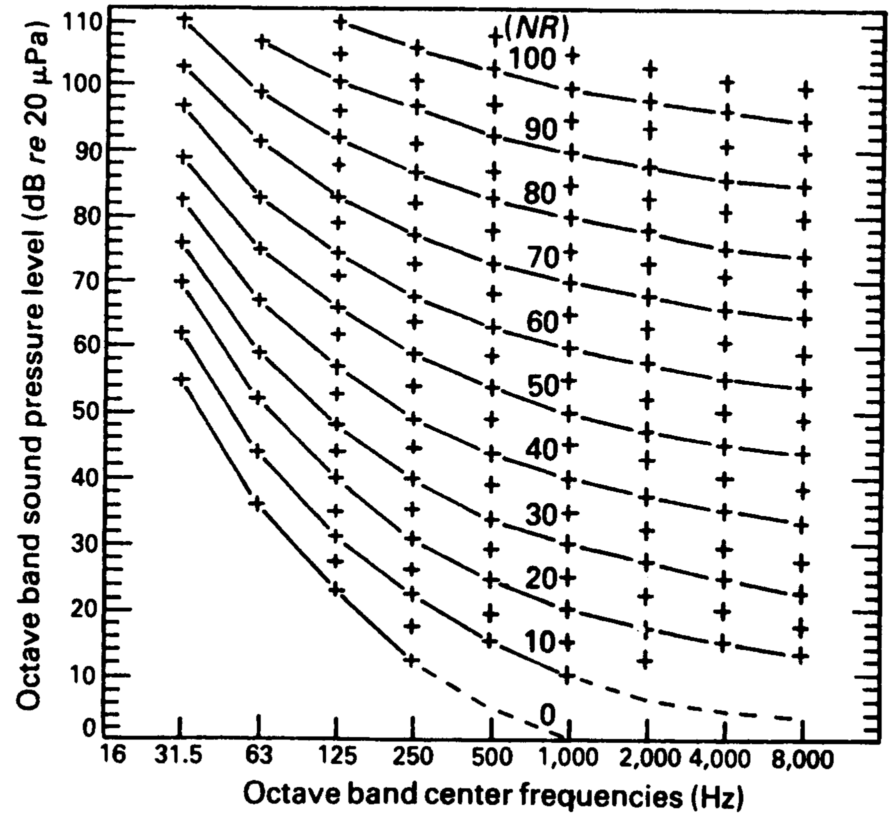


Fig. 8 Noise rating (NR) curves.

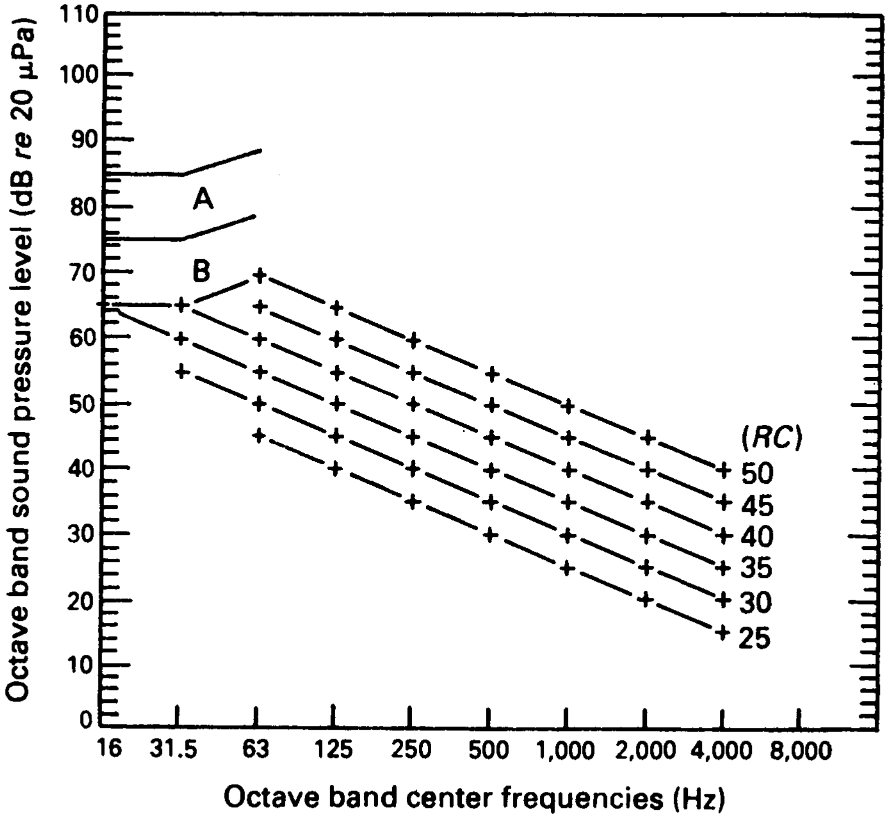


Fig. 9 Room criterion (RC) curves.

that he obtained from the measured data had a slope of about −5 dB/octave, and he thus drew a family of straight lines with this slope (see Fig. 9). He also found that intense low-frequency noise with a level of 75 dB or more in region A is likely to cause mechanical vibrations in lightweight structures (including rattles), while noise in region B has a low probability to cause such vibration. The value of the RC curve is the arithmetic average of the levels at 500, 1000, and 2000 Hz. Since these curves were obtained from measurements made with air-conditioning noise only, they are mostly useful in rating the noise of such systems. RC curves are also discussed in Chapter 94.

8.4 Balanced Noise Criterion Curves

In 1989, Beranek[19,20] modified the NC curves to include the 16- and 31.5-Hz octave bands and changed the slope of the curves so that it is now −3.33 dB/octave between 500 and 8000 Hz. He also incorporated the A and B regions as specified by Blazier in the RC curves. The rating number of a *balanced noise criterion* NCB curve is the arithmetic average of the octave band levels with midfrequencies of 500, 1000, 2000, and 4000 Hz. The result is a set of rating curves that are useful to rate air-conditioning noise in buildings (see Fig. 10). As an example of its use, a background noise spectrum from

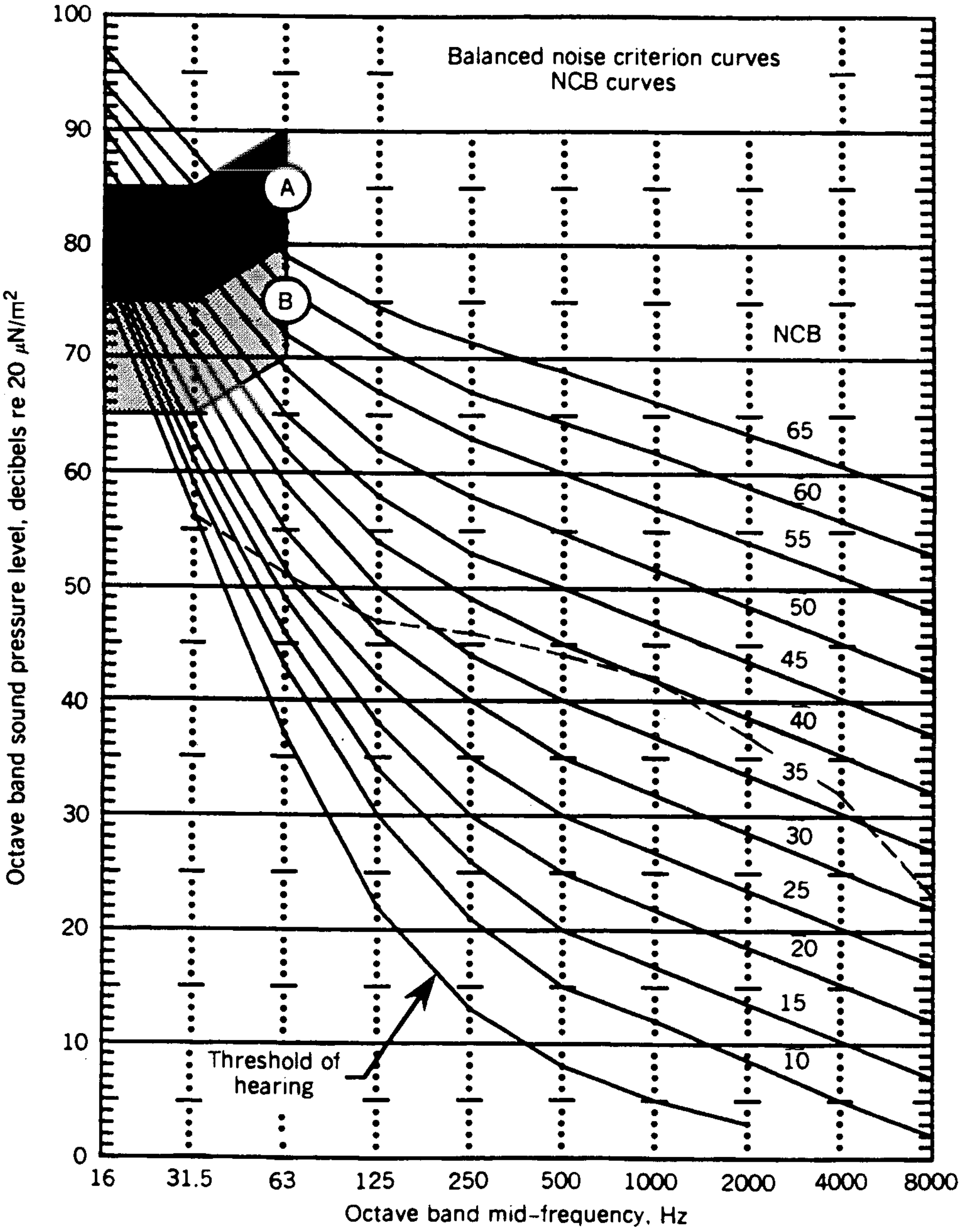


Fig. 10 Balanced noise criterion (NCB) curves.

air-conditioning is plotted in Fig. 8. It is seen that it is tangent to the NCB-40 curve and no octave-band level exceeds this curve. Thus this background noise spectrum can be assigned a rating of NCB 40. Such a noise spectrum might be just acceptable in a general office, and barely acceptable in a bedroom and living room in a house, but not acceptable at all in a church, concert hall, or theater. NCB curves are also described in Chapter 94.

9 EQUIVALENT SOUND LEVEL

For noise that fluctuates in level with time it is useful to define the *equivalent sound level* L_{eq} which is the A-weighted sound pressure level averaged over a suitable period T. This average A-weighted sound level is also sometimes known as the *average sound level* L_{AT}, so that $L_{\mathrm{eq}} = L_{AT}$. The equivalent sound level is given by

$$L_{\mathrm{eq}} = 10 \log \left(\frac{1}{T} \int_0^T p_{\mathrm{A}}^2 dt \Big/ p_{\mathrm{ref}}^2 \right) \quad \mathrm{dB} \qquad (6)$$

where p_{A} is the instantaneous sound pressure measured using an A-weighted frequency filter and p_{ref} is the reference sound pressure 20 μPa. The averaging time T can be specified as desired to range from seconds to minutes or hours.

The average sound level (or the equivalent sound level L_{eq}) can be conveniently measured with an integrating sound level meter or some other similar device. Since it accounts both for magnitude and the duration, L_{eq} has become one of the most widely used measures for evaluating community (environmental) noise from road traffic, railways and industry.[23-25] L_{eq} has also been found to be well correlated with the psychological effects of noise.[26,27] For community noise, a long-period T is usually used (often 24 hours).

10 DAY–NIGHT EQUIVALENT SOUND LEVEL

In the United States during the 1970s, the Environmental Protection Agency developed a measure, from the equivalent sound level, known as the *day–night equivalent level* (DNL) or L_{dn} that accounts for the different response of people to noise during the night.[27]

$$L_{\mathrm{dn}} = 10 \log \frac{15(10^{L_d/10}) + 9(10^{(L_n+10)/10})}{24}, \qquad (7)$$

where L_d is the 15-hour daytime A-weighted equivalent sound level (from 0700 to 2200 hrs) and L_n is the 9-hour nighttime equivalent sound level (from 2200 to 0700 hrs). The nighttime noise level is subjected to a 10 dB penalty because noise at night is deemed to be much more disturbing than noise during the day. This 10-dB nighttime penalty is analogous to the 10-dB nighttime penalty applied in both the composite noise rating CNR and the noise exposure forecast NEF as described in Section 12 which follows. The day–night equivalent level has become increasingly used in the United States and some other countries to evaluate community noise and in particular airport noise. In 1980 the U.S. Federal Interagency Committee on Urban Noise (FICON) adopted L_{dn} as the appropriate descriptor of environmental noise in residential situations.[25,29] Its use is also discussed in Chapter 94, where it is termed day–night average sound level.

11 PERCENTILE SOUND LEVELS

The equivalent sound level discussed above accounts for the fluctuation in noise level of an unsteady noise by forming an average sound pressure level to find an equivalent steady A-weighted sound pressure level. There is, however, some evidence that unsteady noise (e.g., from noise sources such as passing road vehicles or aircraft movements) is more disturbing than steady noise. To try to better account for fluctuations in noise level and the intermittent character of some noises, *percentile sound levels* are used in some measures, in particular those for community and traffic noise.[23,30,31] The level L_n is defined to represent the sound level exceeded n% of the time, and thus L_{10}, for example, represents the sound level exceeded 10% of the time.

Figure 11 gives an example of L_{10}, L_{50}, and L_{90} levels and a cumulative distribution. It is seen in this schematic example figure that the level exceeded 10% of the time L_{10} is 70 dB(A), while L_{50} is sometimes termed the *median* noise level, since for half the time the fluctuating noise level is greater than L_{50} and for the other half it is less. L_{50} is used in Japan for road traffic noise. Levels such as L_1 or L_{10} are used to represent the more intense short-duration noise events. L_{10} is used in Australia and the United Kingdom (over an 18-hour 0600–2400 period) for a target value for new roads and for insulation regulations for new roads. Levels such as L_{90} or L_{99} are often used to represent the minimum noise level, the residual level from a graphic level recorder or the average minimum readings observed when reading a sound level meter. Figure 12 shows the outdoor A-weighted sound levels recorded in 1971 at 18 locations in the United States. Values of $L_{99}, L_{90}, L_{50}, L_{10}$, and L_1 are shown for the period 0700 to 1900 hours.[23,28] The small range in levels in recordings 1 and 4 (urban situations)

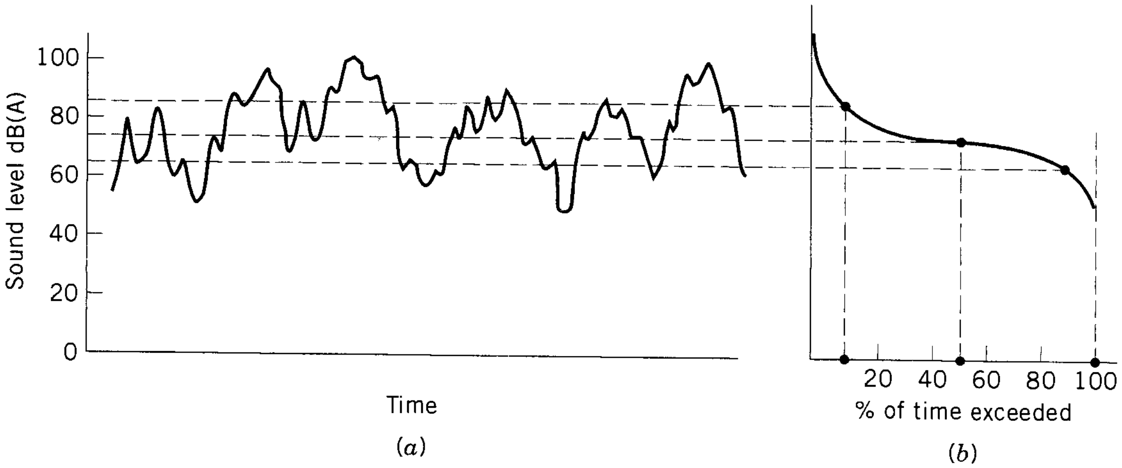


Fig. 11 (*a*) Percentile levels and (*b*) cumulative probability distribution function of percentile levels.

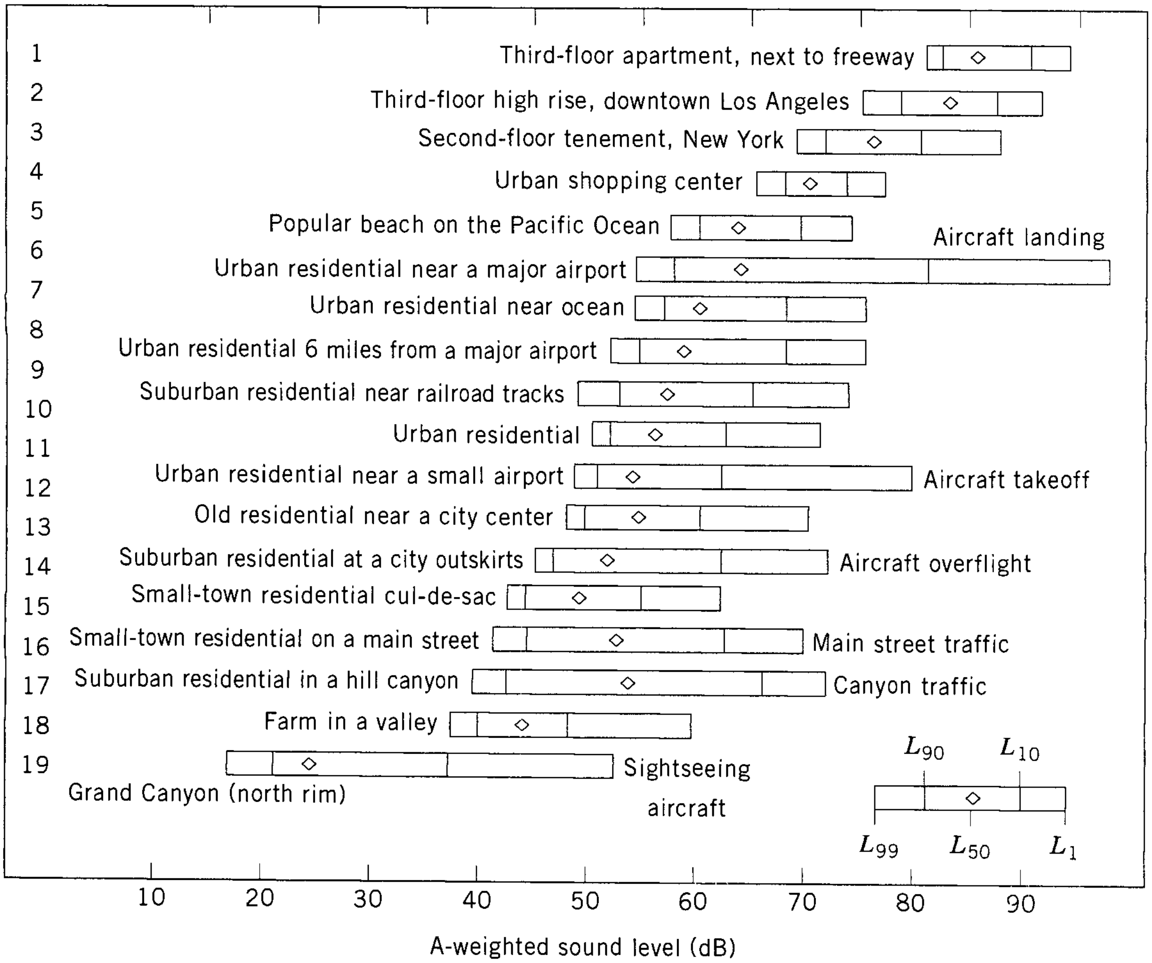


Fig. 12 A-weighted sound levels measured in 1971 at 18 locations in the United States. Values of the percentile levels $L_{99}, L_{90}, L_{50}, L_{10}$, and L_1, where L_N is the A-weighted sound level exceeded N percent of the time are shown for the period 0700–1900 hours. (From Refs. 23, 24, and 28.)

and the large range in levels in recordings 6, 11, 13, and 18 (situations involving aircraft overflights) are very evident. Obviously road traffic usually creates more steady noise, while aircraft movements lead to more extreme variations in noise levels. Percentile sound levels are also discussed in Chapter 94.

12 EVALUATION OF AIRCRAFT NOISE

The noise levels around airports are of serious concern in many countries. Several attempts have been made to produce measures to predict and assess the annoyance caused by aircraft noise in the community. A study of rating measures in 1994 showed 11 different measures in use in the 16 countries studied.[25] The following measures merit brief discussion.

12.1 Composite Noise Rating

The *composite noise rating* (CNR) has a long history dating back to the early 1950s.[32–34] Originally the basic measure it used was the level rank—a set of curves placed about 5 dB apart in the midfrequency range, rather similar to the NC and NR curves described earlier. The level rank was obtained by plotting the noise spectrum on the curves and finding the highest zone into which the spectrum protruded. The rank found initially plus the algebraic addition of corrections gave the CNR. The corrections[32] were for spectrum character, peak factor, repetitive character, level of background noise, time of day, adjustment to exposure, and public relations. The value of CNR obtained was associated with a range of community annoyance categories found from case histories—ranging from no annoyance, through mild annoyance, mild complaints, strong complaints, and threats of legal action, to vigorous community response.

In the late 1950s, the CNR was adapted to apply to the noise of military jet aircraft[35] and later of commercial aircraft.[36] Instead of assigning a level rank, the military aircraft noise was converted to an equivalent sound pressure level (SPL) in the 300–600-Hz range according to a set of curves similar to the level-rank curves.[31,35] The time-varying SPL was time-averaged and then modified by corrections similar to those mentioned above to give the aircraft CNR. The calculation was further modified later when applied to commercial aircraft by using the perceived noise level (PNL) instead of the level rank or the SPL just referred to. The final version of CNR is of the form

$$\text{CNR} = \text{PNL}_{\text{max}} + N + K, \tag{8}$$

where PNL_{max} is the average maximum perceived noise level for individual aircraft flyover events (landing or take-off) for a 24-hour period, N is a correction for the number of aircraft flyover events, and K is an arbitrary constant. The CNR has been found to be useful in predicting where the community reaction to aircraft noise will be on a scale ranging from *no reaction* to *vigorous community response*. The CNR for a certain ground point CNR_{ij} is given by

$$\text{CNR}_{ij} = \text{PNL}_{ij} + 10\log(N_{d,ij} + 16.7N_{n,ij}) - 12, \tag{9}$$

where $N_{d,ij}$ and $N_{n,ij}$ are the number of daytime and nighttime events for each aircraft class i and flight path j. The nighttime flights are penalized by 10 dB. The factor of 16.7 arises because there are fewer nighttime hours than daytime. Then the total CNR for the ground point is given by

$$\text{CNR} = 10\log\sum_i\sum_j \text{antilog}\frac{\text{CNR}_{ij}}{10}. \tag{10}$$

The final version of CNR does not contain any corrections for background noise, previous experience, public relations, or other factors such as the presence of pure tones. Although CNR is no longer used, it is discussed here for completeness and because it has formed the basis for some of the other noise measures and descriptors such as NEF following.

12.2 Noise Exposure Forecast

The *noise exposure forecast* (NEF) is a similar measure to CNR but it uses the effective perceived noise level EPNL instead of PNL.[37,38] Thus NEF automatically takes account of the annoying effects of pure tones and the duration of the flight events. The NEF is of the form

$$\text{NEF} = \text{EPNL} + N + K, \tag{11}$$

where EPNL is the average effective perceived noise level for individual aircraft flyover events over a 24-hour period, N is a correction for the number of flyover events, and K is an arbitrary constant. The NEF at a certain ground point NEF_{ij} is calculated in a similar way to the CNL just described:

$$\text{NEF}_{ij} = \text{EPNL}_{ij} + 10\log(N_{d,ij} + 16.7N_{n,ij}) - 88, \tag{12}$$

where EPNL_{ij} is the effective perceived noise level for the aircraft class i and flight path j and $N_{d,ij}$ and $N_{n,ij}$ are as already defined for CNR. The constant 88 was chosen to make the NEF values distinct from CNR values.

Finally

$$\mathrm{NEF} = 10 \log \sum_i \sum_j \operatorname{antilog} \frac{\mathrm{NEF}_{ij}}{10}. \tag{13}$$

The use of NEF has been superseded in the United States by the day–night level L_{dn}. However, in 1995 NEF remained in use in Canada and in a modified form in Australia.[25]

12.3 Noise and Number Index

The *noise and number index* (NNI) is a subjective measure of aircraft noise annoyance developed and used in the United Kingdom. The NNI was the outcome of a survey in 1961 of noise in the residential districts within ten miles of London (Heathrow) Airport.[39] It is defined by

$$\mathrm{NNI} = \langle \mathrm{PNL} \rangle_N + 15 \log_{10} N - 80, \tag{14}$$

where $\langle \mathrm{PNL} \rangle_N$ is the average peak noise level of all aircraft operating during a day and is given by

$$\langle \mathrm{PNL} \rangle_N = 10 \log \left(\frac{1}{N} \sum_{n=1}^{N} \operatorname{antilog} \frac{\mathrm{PNL}}{10} \right). \tag{15}$$

Here PLN is the peak perceived noise level produced by an individual aircraft during the day and N is the number of aircraft operations over a 24-hour period. Since no annoyance appeared to occur at levels below 80 PNdB, a constant of 80 is used, so that the value of NNI of 0 corresponds to no annoyance.

A second survey[40] in 1967 found that for the same noise exposure the reported annoyance was less than in 1961, though it was not clear whether this was because of noise-sensitive people leaving the area, or noise-insensitive people arriving, or because people adapted or became resigned or apathetic to the noise, or because the background noise from traffic, construction, and industry had increased enough to mask more of the aircraft noise.

In 1988 NNI was superseded in the United Kingdom by a measure based on the A-weighted L_{eq}. The L_{eq} is averaged over the period 0700 to 2300 hours since in the UK nighttime flights are restricted. However, in 1995 NNI remained in use in Ireland and Switzerland.[25]

12.4 Equivalent Sound Level L_{eq} and Day–Night Level L_{dn}

Some countries continue to use NEF or NNI or similar noise measures or descriptors related to these that include a weighting based on the number of aircraft movements.[25] However, because they are much simpler to measure and seem to give adequate correlation with subjective response, there has been a move in several countries towards the use of L_{eq} and L_{dn}.[25] In Germany, Luxembourg, and the United Kingdom L_{eq} has been adopted: (1) in Germany and Luxembourg with day (0600–2200) and night (2200–0600) periods, and (2) in the United Kingdom with an 18-hour period only (0700–2300) (because nighttime flights are normally restricted).

In the United States since publication of the Environmental Protection Agency (EPA) "Levels Document"[27] and other similar publications, the use of CNR and NEF has been largely superseded by the day–night equivalent level DNL for the assessment of the potential impacts of noise and for planning recommendations and land-use management near civilian and military airports. In fact, in 1977 the Office of the Secretary of Defense adopted a new Part 256 of Title 32 of the Code of Federal Regulations, which contains policies for Air Installation Compatible Use Zones (AICUZ). Section 256.3 states that the day–night average sound level DNL will be used to assess the impact of noise with air installations, and Section 256.10 prohibits a further use of the composite noise rating CNR or the noise exposure forecast NEF.

13 EVALUATION OF TRAFFIC NOISE

13.1 The Traffic Noise Index

In an attempt to develop acceptability criteria for traffic noise from roads in residential areas, Griffiths and Langdon[41] produced a unit for rating traffic noise, the *traffic noise index* (TNI):

$$\mathrm{TNI} = 4(L_{10} - L_{90}) + L_{90} - 30. \tag{16}$$

They measured A-weighted traffic noise at 14 sites in the London area and interviewed 1200 people at these sites in the process. Griffiths and Langdon excluded sites with noise sources other than traffic. They then used regression analysis to fit curves to the data. This indicated that L_{10} was better at predicting dissatisfaction than L_{50} or L_{90}, and that TNI was also superior to L_{10}, L_{50}, and L_{90}.

Use of the traffic noise index has not been widespread. The index attempts to make an allowance for the noise variability (with the first term in the above equation) since fluctuating noise is commonly assumed to be more annoying than steady noise. The second term, L_{90}, represents the background noise level, while the third term is simply a constant chosen to yield more convenient numbers.

Some doubt has been cast on the conclusions of Griffiths and Langdon, and it has been suggested that the very short sample times (100 seconds in each hour) used may have resulted in underestimates of L_{10} and overestimates of L_{90}.[42] TNI is not considered today to be significantly superior to either L_{10} or L_{eq} and has not been widely used.

Instead of TNI, the British government has adopted the A-weighted L_{10}, averaged over 18 hours from 0600 to 2400 hrs, as the noise index to be used to implement planning and remedial measures to reduce the impact on people of road traffic noise.[43–45] In addition the British government uses a 16-hour L_{eq} and an 8-hour L_{eq} for the case of land used for residential development.

13.2 Noise Pollution Level

In a later survey, Robinson[46] again concluded that, with fluctuating noise, L_{eq}, the equivalent continuous sound level on an energy basis, was an insufficient descriptor of the annoyance caused by fluctuating noise. He included another term in his *noise pollution level* NPL or L_{NP}, which he defined as

$$L_{NP} = L_{eq} + k\sigma, \qquad (17)$$

where k is a constant and σ is the standard deviation of the sound level. He also offered an alternative expression:

$$L_{NP} = L_{eq} + a(L_{10} - L_{90}). \qquad (18)$$

Robinson examined the available Griffiths and Langdon data[41] and concluded that $a = 1.0$ and $k = 2.56$ were good choices for the constants in these equations. He then examined the aircraft noise experiments of Pearsons[47] and found that L_{NP} predicted very well Pearsons' data points and the trade-off between duration and level for individual flyover events. A-weighted levels were used in L_{NP} with traffic noise, and perceived noise levels (PNdB) were used with aircraft noise. The noise pollution level is potentially attractive because, in principle, it allows annoyance from aircraft, traffic, and perhaps other sources such as industrial noise to be determined. However, the superiority of L_{NP} over all other forms of noise rating has not been proved in practice, and it has not been widely used.

13.3 Equivalent Sound Level

Figure 13 shows the annoyance results of Pearsons and co-workers, using six different noise ratings: L_1, L_{10}, L_{50}, L_{90}, L_{eq}, and L_{NP}.[48] As expected for all of the noise measures, annoyance increases with level. The shapes of the curves, however, do vary considerably when the annoyance is less than very annoying. In particular, the L_{10}

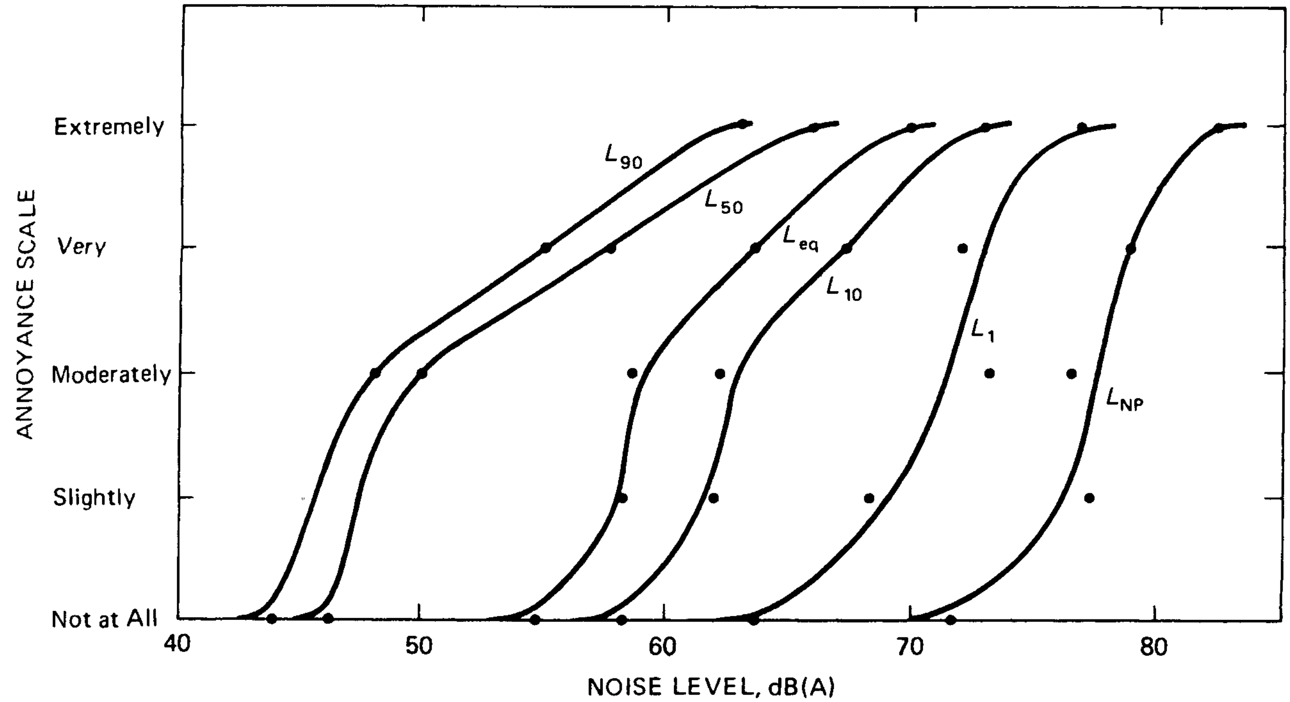


Fig. 13 Annoyance as a function of noise level. (From Ref. 48.)

and L_{eq} measures exhibit a very steep rise in annoyance from the categories of *slightly* to *moderately annoying* for no increase in noise level, presumably one of their drawbacks. However, except for the case of L_{NP} in the *extremely annoying* category, the standard deviation of L_{eq} for a specified response category was in all cases less than or equal to the standard deviation of the other noise measures. This is an advantage in the use of L_{eq} since there is more confidence in the annoyance scores predicted. There is one clear advantage of L_{eq} over L_{10}, however, in the case of noise containing short-duration, high-level single events. If the events do not occur for more than 10% of the time, then L_{10} will be insensitive to these high-level events and will tend to represent the "background" noise. In fact, for a noise measure L_n to be useful, the intruding noise events must be present for *more* than *n*% of the time. This suggests that L_{10} would be unsuitable as a measure of aircraft noise annoyance and that it might be a possible source of error in Griffiths' and Langdon's results[41] for low traffic flows.

The equivalent sound level L_{eq} (often now devoted as L_{Aeq} has become the measure most commonly used to assess and regulate road traffic (and railroad noise).[25] In the United States L_{dn} (a similar measure) is used for road traffic noise assessment.

14 EVALUATION OF COMMUNITY NOISE

In some community noise measures, corrections are applied to community noise levels to account for pure-tone components or impulsive character, seasonal corrections (summer or winter when windows are always closed), type of district (rural, normal suburban, urban residential, noisy urban, very noisy urban) and correction for previous exposure (such corrections are similar to those for NR already discussed).

Figure 14 gives three examples of the A-weighted *sound* levels measured in a community over a 24-hour period.[23] The triangles in the three figures are the maximum levels read from a graphic level recorder. The continuous lines are percentile levels measured for hour-long periods throughout the 24 hours. The highest percentile measure L_1 does not represent the maximum levels well, which are presumably mostly higher-level, short-duration sounds (occurring less than 1% of the time). It is observed in examples (1) and (2) in Fig. 14 that although the day–night levels L_{dn} are only 3 dB different (86 and 83) that there is a much greater fluctuation in sound level with time in example 2. This example is for a location near a major airport and the use of either TNI (Eq. 16) or NPL (Eqs. 17 or 8) would suggest that the noise environment in location 6 would be much more annoying than in location 1 (near a freeway). Example 3 illustrates another quite different distribution of sound levels with time.

In the United States the A-weighted L_{eq} and L_{dn} are normally used to characterize community noise. However, HUD also recognizes the usefulness of L_{10} in some instances. In other countries the A-weighted L_{eq} is mainly used for community noise studies rather than L_{dn}. The international standards on the description and measurement of environmental noise[49] recommend the use of A-weighted L_{eq} and rating levels (which are A-weighted L_{eq} values to which tone and impulse adjust-

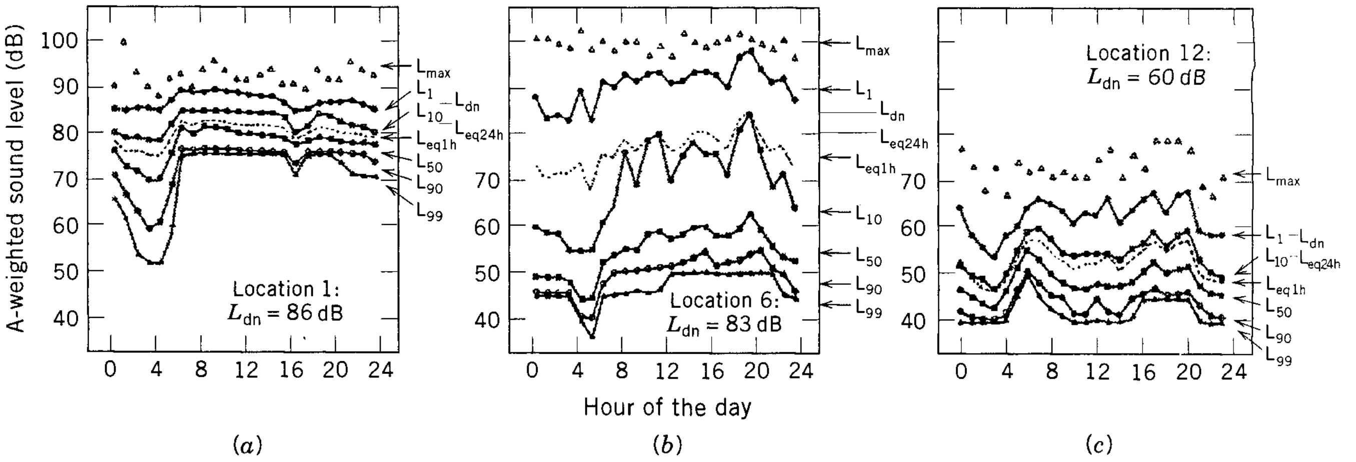


Fig. 14 Variation of percentile levels throughout 24 hour periods recorded at three residential locations (numbers 1, 6, and 12 in Fig. 12) in Los Angeles in 1971, (*a*) third-floor apartment near freeway, (*b*) urban residential location near major airport, and (*c*) old residential location near city center. The values of L_N represent the A-weighted sound levels exceeded *N* percent of the time during one-hour periods. Hourly maximum levels L_{max}, 1-hour and 24-hour values of L_{eq}, and the day–night level L_{dn} are also shown. (From Refs. 23 and 28.)

ments have been made). These standards also recommend that in some circumstances it may be useful to determine the distribution of A-weighted sound pressure levels by determining percentile levels such as L_{95}, L_{50}, and L_5.

In California some state legislation requires the use of the *community noise equivalent level* (CNEL) rather than the day–night level L_{dn}. The two descriptors are similar except that the CNEL has three periods instead of two. Besides the night penalty of 10 dB, an evening penalty of 5 dB is applied with CNEL. CNEL is defined[50] as

$$\text{CNEL} = 10\log_{10}(\textstyle\sum_{0500}^{1900} 10^{(\text{HNLD})/10} + 3 \times 10^{(\text{HNLE}+5)/10} + 9 \times 10^{(\text{HNLN}+10)/10})/24 \qquad (19)$$

where HNLD, HNLE, and HNLN are the day, average evening, and average night hourly noise levels (HNL). The hourly noise level is given by

$$\text{HNL} = 10\log_{10}\frac{\int 10^{L/10}\,dt}{3600}, \qquad (20)$$

where L is the instantaneous A-weighted sound level. The integral is calculated and summed. HNL is usually computed electronically. As is discussed in Chapter 87 the CNEL has also been extensively used to evaluate airport noise in California.[51] In California it has also been used to assess environmental noise transmission into buildings (see Chapter 94).

15 HUMAN RESPONSE

15.1 Sleep Interference

Various investigations have shown that noise disturbs sleep.[52–56] It is well known that there are several stages of sleep and that people progress through these stages as they sleep.[52] Noise can change the progression through the stages and if sufficiently intense can awaken the sleeper. Most studies have been conducted in the laboratory under controlled conditions using brief bursts of noise similar to aircraft fly-overs or the passage of heavy vehicles. However, some have been conducted in the participants' bedrooms. Different measures have been used such as A-weighted *maximum sound level* $L_{A\max}$, A-weighted *sound exposure level* (ASEL) *effective perceived noise level* (EPNdB) and *day–night sound level* (DNL) and most studies have concentrated on the percentage of the subjects that are awakened. Recently Pearsons et al.[57] have reassessed data from 21 sleep disturbance studies (drawn from the reviews of Lucas[52] and Griefhan[53] and seven additional studies). From these data, Finegold et al.[58] have proposed sleep disturbance criteria based on the indoor A-weighted *sound exposure level* (ASEL). In the reanalysis, because of the extremely variable and incomplete data bases, the data were averaged in 5-dB intervals to reduce variability. A regression fit to these data gave the following expression (which is also shown graphically in Fig. 15):

$$\%\text{Awakenings} = 7.1 \times 10^{-6} L_{\text{AE}}^{3.5} \qquad (21)$$

where L_{AE} is the indoor A-weighted sound exposure level ASEL. Although the authors recognize that there are concerns with the existing data and there is a recognition that additional sleep disturbance data are needed, Finegold et al.[58] have proposed that Fig. 15 be used as a practical sleep disturbance curve until further data become available.

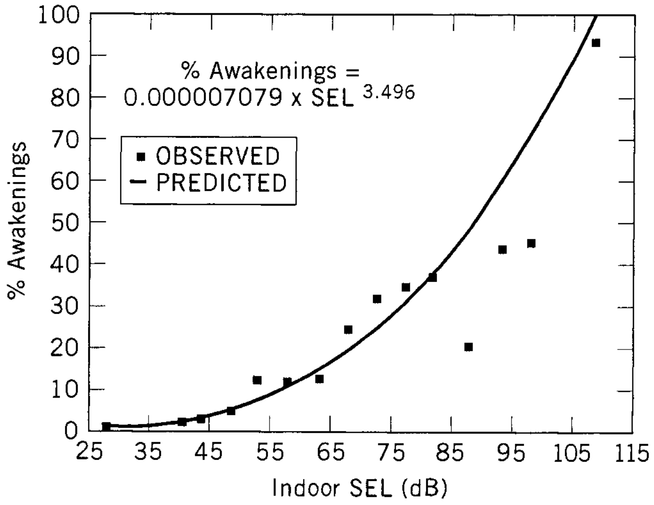


Fig. 15 Proposed sleep disturbance curve based on data of Pearsons et al.[57] (from Ref. 58 with permission). Curve represents percentage of subjects awakened against A-weighted sound exposure level ASEL.

15.2 Annoyance

In 1978 Schultz published an analysis of 12 major social surveys of community annoyance caused by transportation noise.[59] This Schultz analysis which relates the percentage of the population that report they are "highly annoyed" by transportation noise to the day–night average sound level DNL has become widely used all over the world as an important response curve for describing the average community response to environmental noise. Since the Schultz curve was published, additional data have become available. Fidell et al.[60] used 453 exposure response data points compared to the 161 data points originally used by Schultz. This resulted in almost tripling the database for predicting noise annoy-

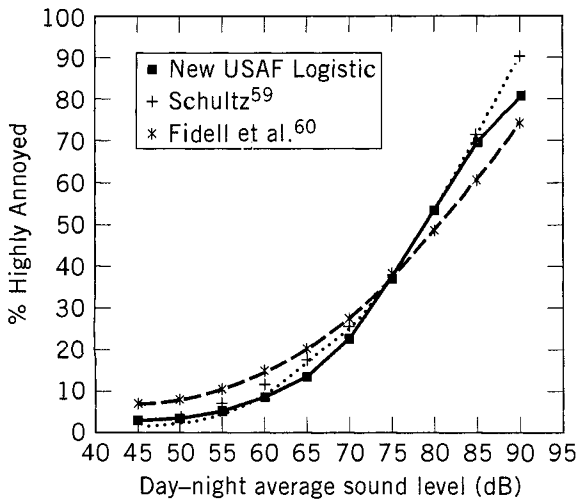


Fig. 16 Curves representing the percentage of subjects that are highly annoyed by noise against A-weighted day–night average sound level DNL: —■— New logistic USAF curve (400 data points), · · · ·+· · · · Schultz[59] third-order polynomial (161) data points) and --*-- Fidell et al.[60] quadratic curve (453 data points). (From Ref. 58 with permission.)

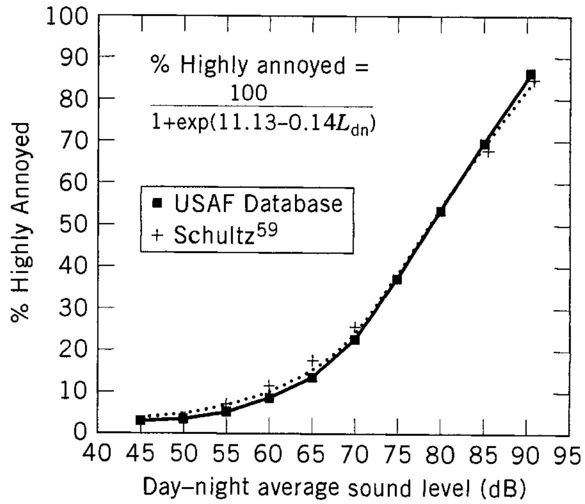


Fig. 17 Curves representing the percentage of subjects that are highly annoyed by noise against A-weighted day–night average sound level DNL. a) —■— Logistic fit to 400 community annoyance social survey data points and b) · · · + · · · 1978 Schultz[59] curve. (From Ref. 58 with permission.)

ance from transportation noise. A later study by the U.S. Air Force eliminated 53 of these data points because there was insufficient correlation between the DNL and the percentage of the population that was highly annoyed %HA.[58] The results of these two studies and the original Schultz curve are given in Fig. 16. Finegold et al.[58] recommend a logistic fit:

$$\%\text{HA} = \frac{100}{1 + \exp(11.13 - 0.14L_{\text{dn}})} \tag{22}$$

rather than the quadratic fit used by Fidell et al.[60], or the third order polynomial fit used by Schultz.[59] This results in a very close agreement between curve obtained with the 400 data points and the original 1978 Schultz curve.[59] See Fig. 17. The differences between the curves in Figs. 16 and 17 are not very significant, however, there are several advantages to the use of the logistic fit given in Eq. (22) including (1) same predictive utility as both the original Schultz curve and the Fidell et al. curve, (2) it allows prediction of annoyance to approach but not reach 0% or 100%, (3) it approaches a 0% community annoyance prediction for a DNL of approximately 40 dB rather than the anomaly of an increase in annoyance for levels of less than 45 dB as predicted by the Fidell et al. curve, (4) use of a logistic function has had a history of success with U.S. federal environmental impact analyses, and (5) it is based on the most defensible social survey database.[58]

Most community noise impact studies since the late 1970s have been based on a combination of aircraft and surface transportation noise sources. However here has been a continuing controversy over whether all types of transportation should be combined into one general curve for predicting community annoyance to transportation noise.[58,61–67] Some researchers have suggested that people find aircraft noise more annoying than traffic noise or railroad noise.[25,58,61,65]

The differences have been discussed in the literature[25,28,58] and can perhaps be explained by a variety of causes such as (1) methodological differences, (2) variability in the criterion for reporting high annoyance, (3) inaccuracy in some of the acoustical measurements, (4) community response biases, (5) aircraft noise entering homes through parts of the building structure with less transmission loss (such as the roof rather than the walls with aircraft noise). Figure 18 shows logistic fits to 400 final data points from a total of 22 different community annoyance surveys. It can be seen that aircraft noise appears to produce somewhat more annoyance than railroad or traffic noise particularly for the higher DNL values. Miedema has made a recent reanalysis of data from selected social surveys which also shows that aircraft noise appears to cause more annoyance than other surface transportation noise sources.[61] However the results from Miedema's study seem to suggest that for high values of DNL (over 60 to 70 dB) that although aircraft noise is by far the most annoying source, that railroad noise is also more annoying than traffic noise (in contrast to the results of Ref. 58.)

If the five causes discussed above can be dismissed as responsible for the apparent greater annoyance of aircraft

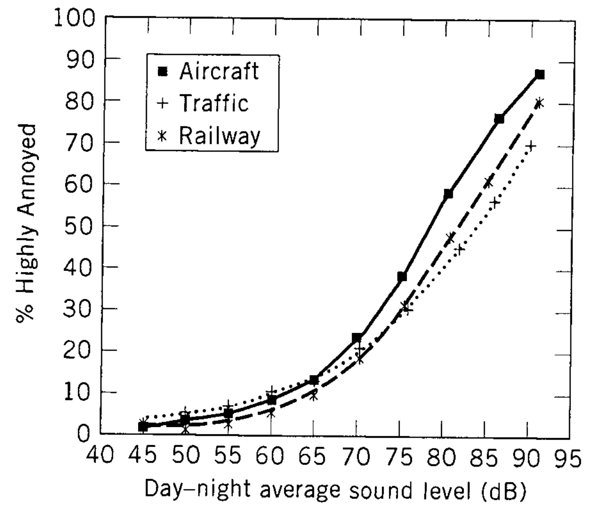


Fig. 18 Curves representing the percentage of subjects that are highly annoyed by noise against day–night average sound level DNL for different sources: —■— aircraft, · · ·+· · · traffic, and --*-- railway. Curves based on data from Fidell et al.[60] (From Ref. 58 with permission.)

noise, then it may be that aircraft noise is more annoying than surface transportation noise for reasons such as the greater variation in level with time and the different frequency spectra from other types of transportation noise sources. As already discussed in the text accompany Fig. 14 in Section 14, aircraft noise does generally exhibit a much greater variation in level from traffic noise and other sources of surface transportation. If such variation is indeed more annoying and is one of the main causes of the difference in annoyance caused by these different forms of transportation noise, this suggests that it may be advisable to re-examine such measures that account for variation in level such as the traffic noise index TNI or the noise pollution level NPL discussed in Sections 13.1 and 13.2.

16 NOISE CRITERIA AND NOISE REGULATIONS

Using some of the noise measures and descriptors discussed and surveys and human response studies, various criteria have been proposed so that noise environments can be determined that are acceptable for people, for speech communication, for different uses of buildings, for sleep and for different land uses. In some countries such criteria are used to write noise regulations for new machinery, vehicles, traffic noise, railroad noise, aircraft and airport noise, community noise, and land use and planning, etc. It is beyond the scope of this chapter to give a comprehensive summary of all these criteria and regulations. Instead just a few are described in this section. The interested reader is referred to the chapters fol-

TABLE 2 Recommended Values of NCB Curves for Different uses of Spaces in Buildings[a]

Type of Space (and Acoustical Requirements)	NCB Curve	Approximate L_A, dB
Broadcast and recording studios	10	18
Concert halls, opera houses, and recital halls	10–15	18–23
Large auditoriums, large drama theaters, and large churches	<20	28
Broadcast, television, and recording studios	<25	33
Small auditoriums, small theaters, small churches, music rehearsal rooms, large meeting and conference rooms	<30	38
Bedrooms, sleeping quarters, hospitals, residences, apartments, hotels, motels, etc.	25–40	38–48
Private or semiprivate offices, small conference rooms, classrooms, and libraries	30–40	38–48
Living rooms and drawing rooms in dwellings	30–40	38–48
Large offices, reception areas, retail shops and stores, cafeterias and restaurants	35–45	43–53
Lobbies, laboratory work spaces, drafting and engineering rooms, and general secretarial areas	40–50	48–58
Light maintenance shops, industrial plant control rooms, office and computer equipment rooms, kitchens and laundries	45–55	53–63
Shops, garages, etc. (for just acceptable speech and telephone communication)	50–60	58–68
For work spaces where speech or telephone communication is not required, but where there must be *no risk* of hearing damage	55–70	63–78

Source: Based in part on Ref. 13.

[a] Also given are the approximate equivalent A-weighted sound levels L_A.

lowing in Part VIII and to Refs. 25, 28, and 62 for more complete summaries of criteria, regulations and legislation. For instance, Chapter 88 provided information on limits for the noise of new vehicles in different countries (where acceleration noise tests are used). Such limits are based on results such as those presented in Fig. 4.

16.1 Noise Criteria

An example of noise criteria is Table 2 which is based on those suggested by Beranek[13] and gives recommended NCB curve values (and approximate A-weighted levels) for various indoor functional activity areas. The NCB curves are given in Figure 10. For example, the air-conditioning unit chosen to supply air to bedrooms (used in residences, apartments, hotels, hospitals, etc.) should have a spectrum corresponding to no more than an NCB curve of 25–40 (or an A-weighted sound level of no more than about 38–48 dB).

Another example of noise criteria are the guidelines recommended by EPA,[27] WHO,[64] FICON,[29] and various European road traffic regulating bodies. See Table 3. As already mentioned L_{eq} is very widely used to evaluate road traffic, railroad, and even aircraft noise[25]. Interestingly, railroad noise has been found to be less annoying than traffic noise in several surveys.[25,68,69] This has resulted in noise limits (using L_{eq}) that are 5 dB lower for railroad noise than road traffic noise (in Austria, Denmark, Germany, and Switzerland) and 3 dB lower in The Netherlands.[25] Gottlob terms this difference the "railway bonus."[25] This seems to contradict the results shown in Fig. 18 and suggests the need for further research.

An example of national noise exposure criteria is the guidance given in the recent British Government guidelines adopted in 1994 for land development given in Table 4. This table shows guidelines in A-weighted sound levels L_{eq} for four noise exposure categories.[70] The noise exposure categories can be interpreted as

TABLE 3 Guidelines from EPA,[27] WHO,[64] FICON,[29] and Various European Agencies for Acceptable Noise Levels

Authority	Specified Sound Levels	Criterion
EPA Levels Document[27]	$L_{dn} \leq 55$ dB (outdoors) $L_{dn} \leq 45$ dB (indoors)	Protection of public health and welfare with adequate margin of safety
WHO Document (1995)[64]	$L_{eq} \leq 50/55$ dB (outdoors; day) $L_{eq} \leq 45$ dB (outdoors; night) $L_{eq} \leq 30$ dB (bedroom) $L_{max} \leq 45$ dB (bedroom)	Recommended guideline values
U.S. Interagency Committee (FICON)[29]	$L_{dn} \leq 65$ dB	Considered generally compatible with residential development
	$65 \leq L_{dn} \leq 70$ dB	Residential use discouraged
Various European road traffic regulations[25]	$L_{eq} \geq 65$ or 70 dB (day)	Remedial measures required

Source: Based on Ref. 28.

TABLE 4 Guidelines Used in the United Kingdom for A-Weighted Equivalent Sound Levels for Different Noise Exposure Categories

		Noise Exposure Category			
Noise Source		A	B	C	D
Road traffic	(07.00–23.00)	<55	55–63	63–72	>72
	(23.00–07.00)	<45	45–57	57–66	>66
Rail traffic	(07.00–23.00)	<55	55–66	66–74	>74
	(23.00–07.00)	<45	45–59	59–66	>66
Air traffic	(07.00–23.00)	<57	57–66	66–72	>72
	(23.00–07.00)	<48	48–57	57–66	>66
Mixed sources	(07.00–23.00)	<55	55–63	63–72	>72
	(23.00–07.00)	<45	45–57	57–66	>66

Source: Based on Ref. 70.

follows[71]: (a) Noise need not be considered as a determining factor in granting planning permission, although the noise level at the high end of the category should not be regarded as a desirable level; (b) noise should be taken into account when determining planning applications, and where appropriate, conditions should be imposed to ensure an adequate level of protection against noise; (c) planning permission should not normally be granted; where it is considered that permission should be given, for example because there are no alternative quieter sites available, conditions should be imposed to ensure a commensurate level of protection against noise; (d) planning permission should normally be refused.

REFERENCES

1. S. S. Stevens, "Procedure for Calculating Loudness: Mark VI," *J. Acoust. Soc. Am.*, Vol. 33, 1961, p. 1577.
2. E. Zwicker, "Ein Verfahren zur Berechnung der Lautstarke (A Means for Calculating Loudness)," *Acustica*, Vol. 10, 1960, p. 304.
3. E. Zwicker, "Subdivision of the Audible Frequency Range into Critical Bands (Frequensgruppen)," *J. Acoust. Soc. Am.*, Vol. 33, 1961, p. 248.
4. Anon., "Method for Calculating Loudness Level," ISO Recommendation R 532-1967, 1967.
5. K. D. Kryter, *The Effects of Noise on Man*, 2nd ed. Academic Press, Orlando, 1985.
6. K. D. Kryter, "Scaling Human Reactions to the Sound from Aircraft," *J. Acoust, Soc. Amer.*, Vol. 31, No. 11, 1959, pp. 1415–1429.
7. K. D. Kryter and K. S. Pearsons, "Some Effects of Spectral Content and Duration on Percieved Noise Level," *J. Acoust Soc. Amer.*, Vol. 35, No. 6, 1963, pp. 866–883.
8. ISO, "Procedure for Describing Aircraft Noise Heard on the Ground," ISO 3891-1978, Int. Std. Org., Geneva, Switzerland, 1978.
9. D. N. May, "Basic Subjective Responses to Noise," in D. N. May (Ed.), *Handbook of Noise Assessment*, Van Nostrand Reinhold, New York, 1978.
10. C. M. Harris (Ed.), *Handbook of Acoustical Measurements and Noise Control*, 3rd ed., McGraw-Hill, New York, 1991, Chapter 47.
11. J. S. Anderson and M. Bratos-Anderson, *Noise, Its Measurement, Analysis, Rating, and Control*, Avebury Technical, Aldershot, UK, 1993.
12. M. J. Crocker, "Noise of Air Transportation to Non-travellers," in D. N. May (Ed.), *Handbook of Noise Assessment*, Van Nostrand Reinhold, New York, 1978.
13. L. L. Beranek and I. L. Ver, *Noise and Vibration Control Engineering*, Wiley, New York, 1992.
14. N. R. French and J. C. Steinberg, "Factors Governing the Intelligibility of Speech Sounds," *J. Acoust. Soc. Am.*, Vol. 19, 1947, pp. 90–119.
15. ANSI, "Methods for the Calculation of the Articulation Index," ANSI S3.5-1969 (R1976), American National Standards Institute, New York, NY, 1976.
16. J. C. Webster, "Communicating in Noise in 1978–1983," in E. Rossi (Ed), *Proceedings of the 4th International Congress on Noise as a Public Health Problem*, Torino, Italy, Volume I, 1983, pp. 411–424.
17. L. L. Beranek, "Revised Criteria for Noise Control in Buildings," *Noise Control*, Vol. 3, 1957, pp. 19–27.
18. L. L. Beranek, W. E. Blazier, and J. J. Figiver, *J. Acoust. Soc. Amer.*, Vol. 50, 1971, pp. 1223–1228.
19. L. L. Beranek, "Balanced Noise Criterion (NCB) Curves," *J. Acoust. Soc. Amer.*, Vol. 86, pp. 650–664.
20. L. L. Beranek, "Applications of NCB Noise Criterion Curves," *Noise Control Engineering Journal*, Vol. 33, 1989, pp. 45–56.
21. W. E. Blazier, "Revised Noise Criteria for Application in the Acoustical Design and Rating of HVAC Systems," *Noise Control Engineering*, Vol. 16, 1981, pp. 64–73.
22. C. W. Kosten and G. J. van Os., "Community Reaction Criteria for External Noise, in the Control of Noise," NPL Symposium No. 12, HMSO, London, 1962, pp. 373–382.
23. EPA, "Community Noise," Report No. NTID 330.3, U.S. Environmental Protection Agency, Washington, DC, 1971.
24. K. M. Eldred, "Assessment of Community Noise," *Noise Control Engineering*, Vol. 3, No. 2, 1974, pp. 88–95.
25. D. Gottlob, "Regulations for Community Noise," *Noise/News International*, Vol. 3, No. 4, 1995, pp. 223–236.
26. EPA, "Public Health and Welfare Criteria for Noise," Report No. 550/9-73-002, U.S. Environmental Protection Agency, Washington, DC, 1973.
27. EPA, "Information on Levels of Environmental Noise Requisite to Protect Public Health and Welfare with an Adequate Margin of Safety," Report No. 550/9-74-004, U.S. Environmental Protection Agency, Washington, DC, 1974.
28. E. A. G. Shaw, "Noise Environments Outdoors and the Effects of Community Noise Exposure," *Noise Control Engineering Journal*, Vol. 44, N. 3, 1996, pp. 109–119.
29. "Guidelines for Considering Noise in Land Use Planning and Control," Federal Interagency Committee on Urban Noise, Document 1981-338-006/8071, U.S. Government Printing Office, Washington, DC, 1980.
30. W. E. Scholes and G. H. Vulkan, *Applied Acoustics*, Vol. 2, 1967, pp. 185–197.
31. T. J. Schultz, *Community Noise Ratings*, 2nd Ed., Applied Science, London, 1982.
32. W. A. Rosenblith, K. N. Stevens, and the staff of Bolt, Beranek, and Newman Inc., "Handbook of Acoustic Noise Control, Vol 2-Noise and Man," *WADC TR-52-204*, Wright Air Development Center, Wright Patterson Air Force Base, Ohio, 1953, pp. 181–200.
33. K. N. Stevens, W. A. Rosenblith, and R. H. Bolt, "A Community's Reaction to Noise: Can It Be Forecast?" *Noise Control*, Vol. 1, 1955, pp. 63–71.

34. W. J. Galloway and D. E. Bishop, "Noise Exposure Forecasts: Evolution, Evaluation, Extensions, and Land Use Interpretations," BBN Report No. 1862 for the FAA/DOT Office of Noise Abatement, Washington, DC, December, 1969.
35. K. N. Stevens, A. C. Pientrasanta, and the staff of Bolt, Beranek, and Newman Inc., "Procedures for Estimating Noise Exposure and Resulting Community Reactions from Air Base Operations," WADC TN-57-10, Wright Air Development Center, Wright Patterson Air Force Base, Ohio, 1957.
36. Bolt, Beranek, and Newman Inc., "Land Use Planning Relating to Aircraft Noise," FAA Technical Report, October, 1964; also issued as Report No. AFM86-5, TM-5-365, NAVDOCKS P-38, Department of Defense, 1964.
37. D. E. Bishop and R. D. Horonjeff, "Procedures for Developing Noise Exposure Forecast Areas for Aircraft Flight Operations," FAA Report DS-67-10, Department of Transportation, Washington, DC, August 1967.
38. D. E. Bishop, "Community Noise Exposure Resulting from Aircraft Operations: Application Guide for Predictive Procedure," AMRL-TR-73-105, November 1974.
39. Committee on the Problem of Noise, "Noise–Final Report," HMSO, London, 1963.
40. "Second Survey of Aircraft Noise Annoyance around London Heathrow Airport," HMSO, London, 1971.
41. I. D. Griffiths and F. J. Langdon, "Subjective Response to Road Traffic Noise," *J. Sound Vib.*, Vol. 8, No. 1, 1968, pp. 16–33.
42. T. J. Schultz, "Some Sources of Error in Community Noise Measurement," *Sound and Vibration*, Vol. 6, No. 2, 1972, pp. 18–27.
43. Department of the Environment, "Motorway Noise and Dwellings," Digest 153, Building Research Establishment, Garston, Watford, England, May 1973.
44. Department of Transport, "Calculation of Road Traffic Noise," HMSO, London, 1988.
45. P. G. Abbott and P. M. Nelson, "The Revision of Calculation of Road Traffic Noise 1988," *Acoustics Bulletin* (Institute of Acoustics), Vol. 14, No. 1, 1989, pp. 4–9.
46. D. W. Robinson, "The Concept of Noise Pollution Level," NPL Aero Report AC 38, National Physical Laboratory, Aerodynamics Division, March 1969.
47. K. S. Pearsons, "The Effects of Duration and Background Level on Perceived Noisiness," Report FAA ADS-78, April 1966.
48. Bolt, Beranek, and Newman, Inc. "Establishment of Standards for Highway Noise Levels (Final Report)," Volume 5, Prepared for Transportation Research Board, National Co-operative Highway Research Program, National Academy of Sciences, NCHRP 3-7/3, November 1974.
49. ISO, "Acoustics—Description and Measurement of Environmental Noise—Part 1: Basic Quantities and Procedures, ISO 1996-11 : 1982; Part 2: Acquisition of Data Pertinent to Land Use ISO 1996-22 : 1987; Part 3: Application to Noise Limits. ISO 1996-3 : 1987," Int. Std. Org., Geneva, Switzerland.
50. "Noise Standards for California Airports," California Administrative Code, Title 4, Subchapter 6.
51. J. Goldstein, "Descriptors of Auditory Magnitude and Methods of Rating Community Noise," in R. J. Peppin and C. W. Rodman (Eds.), *Community Noise*, ASTM, Philadelphia, 1979, pp. 38–72.
52. J. S. Lucas, "Noise and Sleep: A Literature Review and a Proposed Criterion for Assessing Effect," *J. Acoust. Soc. Am.*, Vol. 58, No. 6, 1975, pp. 1232–1242.
53. B. Griefahn, "Research on Noise-disturbed Sleep since 1973," *Proceedings of the Third International Congress on Noise as a Public Health Problem*, ASHA Report No. 10, 1980.
54. C. J. Jones and J. B. Ollerhead, "Aircraft Noise and Sleep Disturbance: A Field Study," in R. Lawrence (Ed.), *Proceedings of EURO-NOISE '92*, London, 1992, pp. 119–127.
55. J. B. Ollerhead et al., "Report of a Field Study of Aircraft Noise and Sleep Disturbance," Civil Aviation Authority, London, England, December 1992.
56. K. S. Pearsons, D. S. Barber, B. G. Tabachnik, and S. Fidell, "Predicting Noise-induced Sleep Disturbance," *J. Acoust. Soc. Am.*, Vol. 97, 1995, pp. 331–338.
57. K. S. Pearsons, D. S. Barber, and B. G. Tabachnik, "Analyses of the Predictability of Noise-induced Sleep Disturbance," Technical Report HSD-TR-89-029, Human Science Division (HSDY/YAH U.S. Air Force Systems Command), Brooks Air Force Base, TX, October 1989.
58. L. S. Finegold, C. S. Harris, and H. E. von Gierke, "Community Annoyance and Sleep Disturbance: Updated Criteria for Assessment of the Impacts of General Transportation Noise on People," *Noise Control Engineering Journal*, Vol. 42, No. 1, 1994, pp. 25–30.
59. T. J. Schultz, "Synthesis of Social Surveys on Noise Annoyance," *J. Acoust. Soc. Am.*, Vol. 64, 1978, pp. 377–405.
60. S. Fidell, D. S. Barber and T. J. Schultz, "Updating a Dosage Effect Relationship for the Prevalence of Annoyance Due to General Transportation Noise," *J. Acoust. Soc. Am.*, Vol. 89, 1991, pp. 221–233.
61. H. Miedema, "Response Functions For Environmental Noise in Residential Areas," Publication No. 92.021, Nederlands Instituut voor Praeventieve Gezondheidszorg TNO, Leiden, The Netherlands, June 1993.
62. H. E. Von Gierke and K. McK. Eldred, "Effects of Noise on People," *Noise/News International*, Vol. 1, No. 2, 1993, pp. 67–89.
63. W. Passchier-Vermeer, "Noise and Health," Publication No. A93/02, Health Council of Netherlands, The Hague, 1993.
64. B. Berglund and T. Lindvall (Eds.), "Community Noise," Document prepared for the World Health Organization, Center for Sensory Research, Stockholm, Sweden, 1995.

65. K. D. Kryter, "Community Annoyance from Aircraft and Ground Vehicle Noise," *J. Acoust. Soc. Am.*, Vol. 72, No. 4, 1982, pp. 1222–1242.

66. T. J. Schultz, "Comments on K. D. Kryter's paper, "Community Annoyance from Aircraft and Ground Vehicle Noise," *J. Acoust. Soc. Am.*, Vol. 72, No. 4, 1982, pp. 1243–1252.

67. K. D. Kryter, "Rebuttal by Karl D. Kryter to Comments by T. J. Schultz," *Journal of the Acoustical Society of America*, Vol. 72, No. 4, 1982, pp. 1253–1257.

68. U. Möhler, "Community Response to Railway Noise: A Review of Social Surveys," *J. Sound Vib.*, Vol. 120, 1988, pp. 321–332.

69. J. Lang, "Schallimmission von Schienenverkerhstrecken. Forschungsarbeiten aus dem Verkehrswesen," Band 23, Vienna, Austria, 1989.

70. British Government Planning Policy Guidance PPG, "Planning and Noise," London, England, September 1994.

71. Private Communication with Rupert Thornley-Taylor, 29 April 1996.

81

HEARING PROTECTION DEVICES

E. H. BERGER AND J. G. CASALI

1 INTRODUCTION

A hearing protection device (HPD) is a personal safety product that is worn to reduce the harmful auditory and/or annoying subjective effects of sound. Hearing protectors are often a method of last resort, when other means such as engineering controls or removal of the person from the noisy environment are not practical or economical.

To a large extent, research and development in hearing protection began during World War II in response to the tremendous hearing loss caused by military weapons. Military and industrial hearing conservation programs and the use of hearing protection followed in the early 1950s, with use proliferating in the early 1970s. Interest was spurred again in the 1980s due to federal regulations mandating use of hearing protection in occupational settings.[1]

This chapter, which begins by describing HPDs and how to dispense and use them, continues with the more technical issues of hearing protector physics, measurement, and performance assessment. It also includes guidance on how to estimate protection and the effects of HPDs on auditory perception and concludes with a discussion of special types of HPDs and relevant standards and regulations.

2 TYPES OF HEARING PROTECTION DEVICES

Hearing protection devices may be broadly categorized into earplugs, which are placed into or at the entrance of the earcanal to form a seal and block sound, earmuffs, which fit over and around the ears (circumaural) to provide an acoustic seal against the head, and helmets, which normally encase the entire head. Although in certain cases acoustical concerns may dictate the selection of a particular type of HPD, normally ergonomic considerations, personal preference, and/or compatibility with other safety gear and job requirements are the deciding factors.

2.1 Earplugs

Earplugs are made from materials such as vinyl, silicone, elastomer formulations, slow-recovery closed-cell foam, spun fiberglass, and cotton–wax combinations. They are available in a wide variety of types and styles, some of which are shown in Fig. 1.

Premolded earplugs are formed from flexible materials into conical, bulbous, or other shapes, often including flanges or sealing rings. They usually are available in a range of sizes to fit most ears and are pushed into place in the earcanal, whereupon a seal is made against the canal walls. By contrast, *formable earplugs*, which are made from materials such as slow-recovery foam, fiberglass, and silicone putty, either are formed prior to insertion into the earcanal (e.g., foam plugs are rolled into thin cylinders and reexpand in the canal after insertion) or are pressed into the canal and forced to (de)form to seal the canal at its entrance. *Custom-molded earplugs* are manufactured from an individual impression of the earcanal, which must precisely match the shape of the canal to create a seal and block the noise.

A final type of earplug is the *semi-insert* (also called semiaural, canal caps, or concha seated), which consists of soft pods that are held in place against and/or slightly inside the rim of the earcanal by a lightweight band. They are easily hung around the neck when not in use.

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker
Copyright © 1997 John Wiley & Sons, Inc.

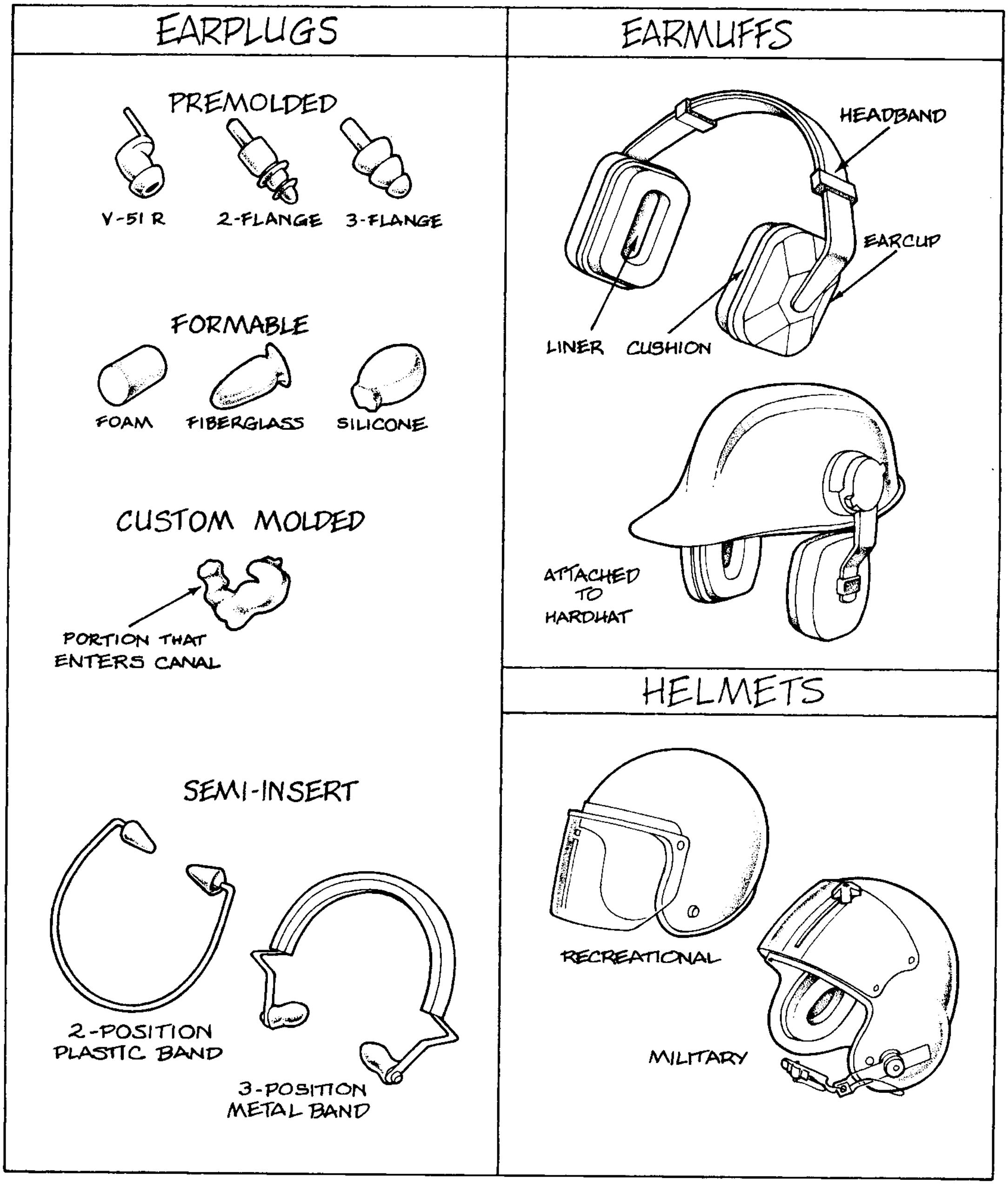

EARPLUGS
EARMUFFS
PREMOLDED
V-51 R
2-FLANGE
3-FLANGE
FORMABLE
FOAM
FIBERGLASS
SILICONE
CUSTOM MOLDED
PORTION THAT ENTERS CANAL
SEMI-INSERT
2-POSITION PLASTIC BAND
3-POSITION METAL BAND
HEADBAND
EARCUP
LINER
CUSHION
ATTACHED TO HARDHAT
HELMETS
RECREATIONAL
MILITARY

Fig. 1 Personal hearing protection devices. From C. W. Nixon and E. H. Berger, "Hearing Protection Devices," in C. M. Harris (Ed.), *Handbook of Acoustical Measurements and Noise Control*, McGraw-Hill, New York, 1991, pp. 21.1–21.24.

2.2 Earmuffs

Earmuffs normally consist of rigid molded plastic earcups that completely enclose and seal around the outer ear (also called the pinna) using foam- or fluid-filled cushions. The earcups are held in place by an adjustable headband or by short spring-loaded arms attached to a hard hat or other headgear. Headbands, which may be of plastic or metal construction, may function in only a single position or be "universal," suitable for use over the head, behind the neck, or under the chin. The earcups are lined with acoustical material, typically foam, to absorb high-frequency (>2 kHz) energy within the cup.

2.3 Helmets

Helmets enclose a substantial portion of the head and are usually designed primarily for impact protection. When they contain circumaural earcups or a dense liner to seal around the ears, they can also provide beneficial amounts of hearing protection.

3 DISPENSING, FITTING, USE, AND CARE

3.1 Initial Dispensing or Purchase

Initial dispensing of HPDs within an occupational hearing conservation program, individually or in small groups, can have a substantial positive effect on the efficacy of the overall hearing conservation effort.[2]

With the aid of a focused light source (such as an otoscope, earlight, or penlight) a visual inspection of the external ear and head should be made to identify medical or anatomical conditions that might interfere with or be aggravated by the hearing protector (e.g., inflammation, tenderness, excessive or impacted cerumen). When such conditions are present, medical consultation and/or corrective treatment should be obtained.[3]

When consumers purchase hearing protectors for occasional nonoccupational use, it is usually neither feasible nor necessary to have an ear examination prior to purchase or use. However, if ear pain or other aural symptoms develop or discomfort persists, use should be discontinued until audiological or medical advice is obtained. Other brands or types of HPDs may need to be evaluated until the one best suited to the individual is found.

3.2 Fitting Tips

Earplugs generally require more skill and attention than earmuffs during the initial issue, but a common mistake is to presume that earmuffs are foolproof and to dispense them indiscriminately, without providing assistance.

When possible, wearers should be exposed to noise during the fitting procedure so they can hear the noise diminish to the lowest perceived level while adjusting the HPD. A portable cassette player can be used to present recordings of broadband noise or representative industrial sounds.

For all types of earplugs, insertion is easier when the pinna is pulled *outward and upward* with the hand opposite the ear being fitted. Earplugs should be inserted into the right ear using the right hand and into the left ear using the left hand. This allows the hand inserting the earplug to have the best line of approach to the earcanal.

A properly inserted premolded earplug will generally create a blocked-up feeling due to the requisite airtight seal. When the seal is present, suction should be felt if an attempt is made to withdraw the plug from the canal.

Foam earplugs are prepared for insertion by rolling them into a *very thin* crease-free cylinder. This is accomplished by squeezing lightly as one begins rolling and then applying progressively greater pressure as the plug becomes more tightly compressed. Unlike other types of earplugs, foam earplugs should not be readjusted while in the ear. If the initial fit is unacceptable, they should be removed, rerolled, and reinserted.

3.3 Occlusion Effect and Its Use as a Test for Fit

The occlusion effect is the increase in efficiency with which bone-conducted sound is transmitted to the ear at frequencies below 2000 Hz when the earcanal is occluded and sealed. This effect causes wearers of HPDs to perceive a change in the loudness and quality of their own voices and other body-generated sounds/vibrations such as those caused by chewing, breathing, and walking.

The effect is easily demonstrated by plugging one's ears with the fingers while reading aloud. The voice will appear to have added fullness, hollowness, or resonant bassiness. The occlusion effect is often cited as an objectionable characteristic of wearing hearing protection. Its magnitude varies with the way in which the ear is occluded, and may be reduced by adjusting earplugs toward a deeper insertion or selecting earmuffs with larger volume earcups. The maximum effect occurs for semi-inserts. Listening for the presence (*not* the magnitude) of the occlusion effect is a useful technique for fitting nearly all types of HPDs, since it indicates the presence of an acoustical seal and proper fit.[4]

3.4 Comfort

Comfort is a critical feature of a hearing protection device, usually equivalent in importance to the attenuation that the device can provide. Although comfort assessments are, of necessity, subjective in nature, research in recent years has been directed at quantifying the important parameters.[5]

An HPD that is uncomfortable will not be worn consistently and correctly or, worse, will not be worn at all. Comfort can be improved by properly matching the device to the wearer and by providing instruction in its proper use. Common sources of discomfort include improperly sized or inserted earplugs, overly tight earmuff headbands, worn-out devices with hardened/cracked cushions or stiffened/ragged flanges, and the use of earmuffs in hot environments.

3.5 Hygiene, Maintenance, and Safety

Hearing protection devices should be cleaned regularly according to manufacturers' instructions. Normally, warm water and mild soap are satisfactory cleansing agents. Earplugs should be washed and dried thoroughly before reuse or storage. Earmuff cushions should be periodically wiped or washed clean. When they cannot be adequately cleaned or no longer retain their original appearance or resiliency, earmuff cushions and earplugs should be replaced.

Although the likelihood of hearing protection devices increasing the incidence of outer ear infections is minimal, such ear infections may occur.[3] Earplugs or earmuffs that may be implicated are commonly found to be contaminated with caustic or irritating substances or embedded with abrasive materials.

The headbands on earmuffs and semi-insert earplugs may lose their force with time or be purposely sprung to reduce force and increase comfort. This will degrade attenuation. Regular inspection of HPDs and proper instructions to the users are necessary to overcome this problem.

Commercially available HPDs are safe to wear as long as the manufacturers' recommendations are followed; however, some precautions are advisable. For example, earplugs that create an airtight seal, such as premolded inserts, should be removed with a slow twisting motion to gradually break the seal as they are withdrawn, to assure that removal does not cause pain or harm to the ear. When selecting earmuffs, smaller, lighter weight (lower profile), dielectric versions will normally be less likely to interfere with workers' activities or create hazards.

4 SOUND PATHWAYS TO THE OCCLUDED EAR

In the *un*occluded ear, the dominant sound path for external sounds is along the earcanal to the eardrum. However, in the occluded ear four distinct and important paths by which sound can penetrate the HPD can be identified, as described below.[4]

4.1 Air Leaks

For maximum protection, HPDs must make virtually an airtight seal with the walls of the earcanal or the circumaural regions surrounding the pinna. Air leakage paths can reduce attenuation by 5–15 dB over a broad range of frequencies depending upon the size of the leak. Typically the loss is most noticeable at low frequencies.

4.2 Hearing Protector Vibration

Due to the flexibility of the earcanal flesh, earplugs can vibrate in a pistonlike manner, thus limiting their low-frequency attenuation. Earmuffs, too, vibrate as a mass–spring system, the stiffness of the spring depending upon the dynamic characteristics of the earmuff cushion and the circumaural flesh as well as the volume of the air entrapped inside the earcup. These actions limit low-frequency attenuation. Representative maximum values at 125 Hz for earmuffs, premolded earplugs, and foam earplugs are 25, 30, and 40 dB, respectively.

4.3 Transmission through Materials

In much the same way as sound will propagate through any barrier, so too will it pass through an HPD, albeit with diminished intensity. The transmission depends upon the mass, stiffness, and internal damping of the HPD materials. The limitations caused by this transmission path are most pronounced above 1 kHz.

4.4 Bone and Tissue Conduction

Even if an HPD were completely effective in blocking the preceding pathways, the sound would still reach the inner ear by bone and tissue conduction, thus imposing a limit on the maximum real-ear attenuation that a device can provide. However, since that limit is 40–50 dB below the sound level reaching the ear through the open earcanal, it does not become important unless attenuation of the HPD itself approaches such values (see Fig. 2). This is not often the case unless dual HPDs are worn (see Section 6.3).

5 MEASURING ATTENUATION

More than a dozen different methods of measuring HPD attenuation have been described in the literature, but only a few have been found practical and reliable and are commonly utilized.[6]

Historically, standards committees have been primarily concerned with the precision of HPD attenuation measurements, that is, the repeatability of the data, both within and between laboratories. Although this is certainly an important issue, experience has shown that undue emphasis on precision, with a consequent lack of regard for accuracy or "realism," may lead to data of questionable utility; by definition, "useful" data predict with reasonable accuracy the protection that can typically, or even optimally, be expected to be achieved in practice.[7]

5.1 Laboratory Procedures

Hearing protector research is usually conducted under laboratory conditions. This allows the experimenter to exert the most control over the relevant parameters of

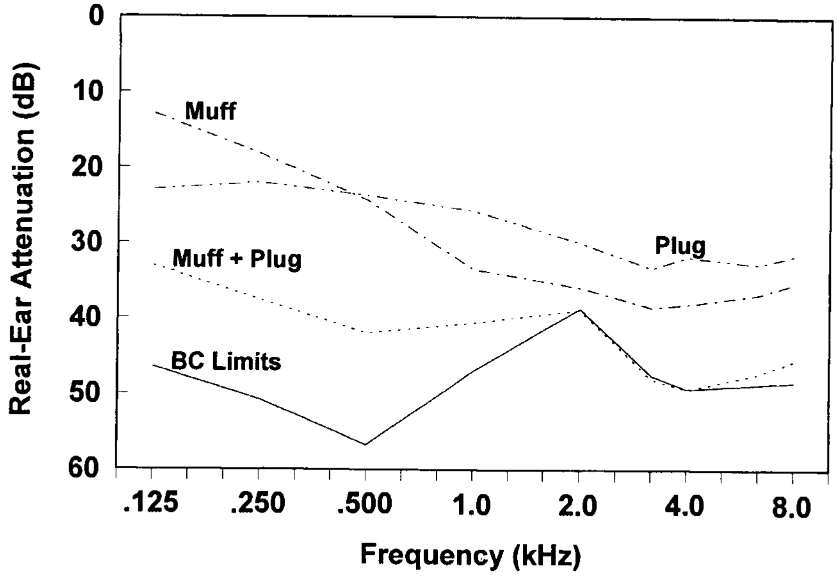


Fig. 2 Bone conduction (BC) limits to hearing protector attenuation and an example of attenuation provided by an earplug, an earmuff, and the two devices worn together.

which he or she is aware and provides the best assurance of generating repeatable test data. However, even the most accurate of laboratory methods will only provide data representative of field (real-world) protection if (a) the HPD is fit and worn by the subjects in the same manner as by wearers in practice and (b) the test population is representative of the actual users.

Real-Ear Attenuation at Threshold The most common and one of the most accurate HPD attenuation tests is the measurement of real-ear attenuation at threshold (REAT).[8,9] Virtually all available manufacturers' reported data have been derived via this method, and it is the procedure required by the Environmental Protection Agency (EPA) to obtain data for computation of the Noise Reduction Rating (NRR).

The REAT is a measurement of the difference between the minimum level of sound that a subject can hear without wearing an HPD (open threshold) and the level needed when the HPD is worn (occluded threshold). The difference is a measure of the real-ear attenuation provided by the device. This method accurately represents the performance of all conventional HPDs (those that provide constant attenuation regardless of sound level, i.e., level-*in*dependent HPDs). For those HPDs whose attenuation increases with level (i.e., amplitude-sensitive HPDs), REAT usually provides a lower bound estimate.

Microphone in Real Ear The microphone-in-real-ear (MIRE) method utilizes physical (acoustical) measurements to determine the difference, termed insertion loss, between the sound pressure levels in the human ear with and without the HPD in place. The levels are measured via a miniature or probe microphone at the entrance to or sometimes in the earcanal itself.[10] Alternatively, a noise reduction measurement utilizes two microphones to simultaneously measure the levels inside and outside the HPD. The subject who wears the HPD is required to sit still and behave as an inanimate acoustic test fixture.

This method can be used to measure attenuation for all types of circumaural HPDs including those fitted with electronics; difficulties can arise with earplugs because of the problems of inserting both a microphone and an earplug simultaneously into the earcanal. The MIRE method is much less time consuming than REAT testing as well as an excellent alternative, especially when measurements are required at high sound levels such as for investigation of the response of HPDs to weapons fire and explosions.

Acoustical Test Fixture The acoustical test fixture (ATF) method is directly analogous to the MIRE method, except that an inanimate fixture is utilized.[10] The ATF incorporates a microphone at the approximate eardrum location and, depending upon its sophistication, may match the impedance of the eardrum (via a suitable acoustic coupler) and also include simulations of the circumaural and earcanal tissues.

The ATF measurements provide a quick assessment for product development, quality control, and acceptance testing of earmuffs and helmets. In general, ATF measurements of earplugs are still ill defined, and for all devices the data are generally not suitable for direct prediction of real-ear attenuation values.

5.2 Field Procedures

Generally, the most accurate laboratory procedures, modified to reduce cost and improve portability and ruggedness, are those utilized under field conditions. The purpose is to derive more accurate estimates of the attenuation provided by HPDs in use. The procedures that have been most successfully utilized are sound-field REAT in a manner similar to that described in the existing standards,[8,9] REAT conducted using large circumaural cups for the stimulus presentation instead of loudspeakers in the test room, and the MIRE approach with microphones simultaneously mounted on the inside and outside of earmuffs.[6]

When the REAT procedures are implemented, workers are usually selected without warning from their place of work and accompanied to a test site, with care taken to assure they do not readjust their HPDs prior to testing. For the MIRE procedure, the instrumented earmuffs are utilized to simultaneously measure the protected and unprotected noise doses received by the exposed workers.

6 ATTENUATION CHARACTERISTICS OF HEARING PROTECTION DEVICES

Especially for insert-type devices, differences among wearer or test subject fitting procedures lead to a wide variation in reported attenuation,[11] even in the case of purportedly "optimum-fit" data measured in the laboratory.[4] Comparison of published NRRs and rank ordering of HPDs is strongly influenced by the particular data available for comparison purposes. Sources of interlaboratory variability are (1) differences in interpretation and implementation of the standardized measurement methodology, (2) uncertainty of obtaining the proper fit to avoid acoustic leaks, (3) differences in subject selection and training, and (4) differences in data reduction techniques. The repeatability of attenuation measurements in the same laboratory is also subject to variation, but to a lesser extent than between laboratories.

Differences in the labeled NRRs of competing products of less than 3 dB have *no practical importance*, as a consequence of the inter- and intralaboratory variability discussed above, and even 4–5-dB differences are of questionable significance unless data from closely controlled tests are being compared.

6.1 Optimum Laboratory Performance

The values reported in this section are representative of those obtained in laboratory-based REAT measurements, wherein the HPDs are fitted either by or under the close supervision of the experimenter.

Earplugs Attenuation values are similar for various types of premolded earplugs, clustering closely around 25 dB at frequencies up to 1000 Hz, and increasing to about 40 dB at the higher frequencies. Formable cotton–wax or silicone earplugs provide a wider range of attenuation across brands or types, with values as low as 10 dB for devices with the least attenuation and ranging from 30 to 45 dB above 2000 Hz.

Formable slow-recovery foam earplugs provide among the highest overall protection of any single device. Attenuation can range from 30 to 45 dB at and above 2000 Hz and, depending upon the depth of insertion, from about 20 to 40 dB below that frequency.

Attenuation of custom-molded earplugs varies widely due to impression-taking and manufacturing procedures as well as differences in materials. On average, the attenuation of these devices tends to be less than found for premolded earplugs.

Generally, earplugs of the premolded and foam varieties provide better attenuation than earmuffs below 500 Hz and equivalent or greater protection above 2000 Hz. At the intermediate frequencies of 500 and 1000 Hz, earmuffs usually provide the highest attenuation.

Semi-insert devices display a wide range of performance at all frequencies. The best devices provide an average attenuation of about 20 dB at frequencies below and 35 dB above 2000 Hz.

Earmuffs The attenuation of earmuffs is substantially controlled by earcup volume and mass, the area of the cup opening, headband force, and the size, shape, and material of the cushion. Attenuation usually increases above 9 dB/octave from 125 to 1000 Hz and, for most earmuffs, at 2000 Hz, approaches the limit imposed by bone conduction of approximately 40 dB. Above that frequency, attenuation averages around 35 dB.

When eyeglasses are worn with earmuffs, the temples should fit close to the sides of the head and be as thin as practical to reduce their effect on the ability of the cushions to seal around the ears. The loss in attenuation that eyeglasses create, with cushions in good condition, is normally 3 to 7 dB, but the effect varies widely among earmuffs and eyeglasses.

Helmets Unless helmets seal well around the head and/or contain internal circumaural cups, the attenuation they provide is very low, especially at the frequencies below 2000 Hz. For example, typical motorcycle helmets provide less than 10 dB of protection below 1000 Hz and only from 10 to 30 dB at the higher frequencies. Some military helmets (e.g., flight helmets), which possess a more effective acoustical design, yield attenuation that is comparable to conventional earmuffs, but still about 5–10 dB less at the frequencies below 2000 Hz. From 4 to 8 kHz, military helmets can surpass earmuff attenuation by about 5 dB if they fully enclose the head, since this decreases bone conduction transmission at high frequencies.

6.2 Laboratory versus Real-World Attenuation

Real-world HPD attenuation is typically much less than found in the laboratory wherein measurements usually reflect the maximum performance that can be expected with a device that is worn under ideal circumstances. Such conditions rarely if ever are duplicated in the workplace. This is partly because of the emphasis in most laboratory testing on obtaining maximum attenuation, disregard of comfort during fitting of protectors (primarily earplugs), and the short time periods that the test subjects wear the devices (5–10 min). But the disparity is also due to the fact that workplace performance usually falls short of what could reasonably be expected to be achieved, because of various deficiencies in hearing conservation programs.

Numerous studies have been conducted in the field to assess real-world performance. The data in the next section are drawn from 22 such studies spanning seven countries, greater than 90 different industrial sites, and approximately 2900 worker/subjects.[12] The results provide an indication of the performance of HPDs, circa 1980–1990, that can be anticipated in the better industrial and military hearing conservation programs.

Comparative Octave Band Data As shown in Fig. 3, attenuation measured in the workplace with actual workers as subjects shows lower mean attenuation values (most notably for earplugs) and higher standard deviations (by a factor of 2–3) than do manufacturers' reported laboratory data. Earmuff performance in the workplace is better relative to earplugs than would be predicted from laboratory data,[13] but regardless of the HPD that is selected, individual noise exposure can be substantially underestimated when it is computed from laboratory attenuation values.

Comparative NRRs Whether attenuation is estimated using laboratory octave band data or via a single-number rating such as the NRR that is computed therefrom (see Section 7.2), the discrepancies with respect to field data are the same.[12] The field NRRs attained by 84% of wearers are less than or equal to 7 dB for all types of earplugs, except the foam variety, which yields about 13 dB, a value roughly comparable to that of most earmuffs.

Attenuation versus Use Time The effective protection of a device (based upon a user's total accumulated noise dose) is substantially reduced when it is removed by an employee for even short periods of time during the workday, as illustrated in Fig. 4. An HPD with a nominal NRR of 25 dB that is removed by the user

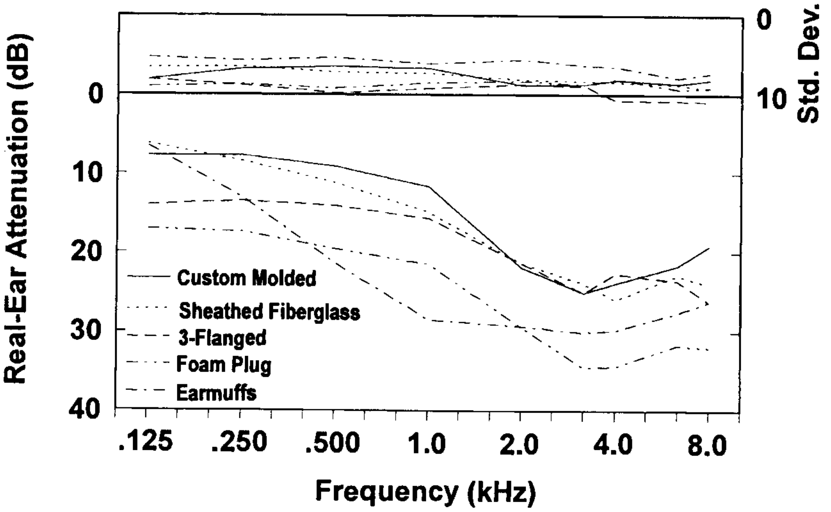


Fig. 3 Summary of real-world data for hearing protectors separated into four earplug and one earmuff categories (from Ref. 12.)

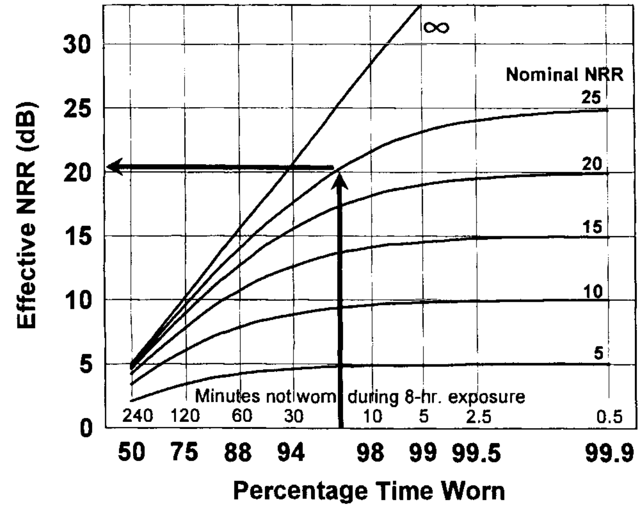


Fig. 4 Corrections to nominal Noise Reduction Ratings (NRRs) as a function of the time during a work day that a hearing protection device is not worn based on a 5-dB trading relationship.

for 15 min during an 8-h workday provides an effective NRR of only 20 dB, for a loss of 5 dB. Greater reductions in effectiveness are observed for higher NRRs.

6.3 Dual Protection

Dual protection, such as earplugs worn in combination with earmuffs, helmets, or communications headsets, typically provides greater attenuation than either device alone.[14] However, the attenuation of the combination is *not* equal to the sum of the separate attenuations, as illustrated in Fig. 2. At individual frequencies the incremental gain in performance varies from approximately 0 to 15 dB over the better single device, but at 2000 Hz the gain is limited to only a few decibels. Attenuation changes very little when different earmuffs are used with the same earplug, but for a given earmuff the choice of earplug is critical for attenuation at frequencies below 2000 Hz. At and above 2000 Hz, essentially all dual-protection combinations provide attenuation approximately equal to the limitations imposed by the bone conduction pathways, approximately 40 to 50 dB, depending upon frequency.

6.4 Infrasound/Ultrasound

At infrasonic frequencies (i.e., below about 20 Hz), well-fitted earplugs provide attenuation that is about equal to that in the 125-Hz one-third-octave band. However, earmuffs provide very little protection from infrasound and may even cause it to be amplified.

At ultrasonic frequencies (i.e., above about 20,000 Hz), conventional earplugs and earmuffs generally provide adequate protection, with attenuation exceeding 30 dB for frequencies from about 10,000–30,000 Hz.

6.5 Weapons Noise

Higher attenuation earmuffs and well-seated earplugs usually provide adequate attenuation of small-arms fire (impulses composed mainly of middle- and high-frequency energy from pistols and standard rifles), reducing the peak sound pressure levels to less than 140 dB. Earmuff attenuation decreases as the peak energy in the impulse moves to lower frequencies, as is the case for larger weapons. The attenuation of peak sound levels provided by a typical earmuff is about 30 dB for pistol fire, about 18 dB for rifle fire, and as little as 5 dB for cannons. Foam earplugs provide attenuation similar to that of earmuffs against pistol fire but are substantially more effective than earmuffs against cannons and bazookas, providing protection greater than 15 dB.[15]

The use of dual HPDs can afford extra protection, especially for low-frequency impulses that have high peak values. Attenuation values of 20–25 dB can be achieved.

7 USING ATTENUATION DATA TO ESTIMATE PROTECTION

The accuracy with which the laboratory or real-world data can predict the protection for an individual, or group of users, is critically dependent upon the correspondence between the test and actual usage conditions. The accuracy can be very poor, as is clearly shown when optimum laboratory data are used to estimate protection in a typical hearing conservation program (see Section 6.2). Undue emphasis on protection estimates is unwarranted.

7.1 Octave Band Method

Potentially the most accurate computational procedure for applying test data to estimate protected exposures is the octave band (OB) method, as illustrated in Table 1. At each frequency, the HPD's mean attenuation less a standard deviation (SD) correction is subtracted from the measured A-weighted OB sound pressure levels. Two SDs are normally used in the United States; one SD is used in Europe and Australia. The protected levels are then logarithmically summed to determine the A-weighted sound level under the HPD. This computation requires that the user have OB noise data available and that the computations be made individually for each HPD–noise spectrum combination.

A critical conceptual error that is often made is to pre-

TABLE 1 Octave Band Calculation of HPD Noise Reduction for a Known Noise Environment[a]

Steps in Procedure	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz	dB (A)[b]
1. Measured sound pressure levels	85.0	87.0	90.0	90.0	85.0	82.0	80.0	
2. A-weighting correction	−16.1	−8.6	−3.2	0.0	+1.2	+1.0	+1.1	
3. A-weighted sound levels (step 1 + step 2)	68.9	78.4	86.8	90.0	86.2	83.0	78.9	93.5
4. Typical premolded earplug attenuation	27.4	26.6	27.5	27.0	32.0	46.0[c]	44.2[d]	
5. Standard deviation × 2	7.8	8.4	9.4	6.8	8.8	7.3[c]	12.8[d]	
6. Estimated protected A-weighted sound levels (step 3 − step 4 + step 5)	49.3	60.2	68.7	69.8	63.8	44.3	47.5	73.0

[a] The estimated protection for 98% of the users in the noise environment assuming they wear the device in the same manner as did the test subjects and assuming they are accurately represented by the test subjects is 93.5 − 73.0 = 20.5 dB (A).

[b] Logarithmic sum of seven octave band levels in the row.

[c] Arithmetic average of 3150- and 4000-Hz data.

[d] Arithmetic average of 6300- and 8000-Hz data.

sume that the SD correction adjusts laboratory data to estimate real-world values. The actual purpose of the SD is to adjust the mean test data to reflect the attenuation to be expected in 84% (for a one-SD correction) or 98% (for a two-SD correction) of the test subjects.

7.2 Noise Reduction Rating

A single-number descriptor is convenient and often sufficiently accurate to estimate protected exposures. One such descriptor, precalculated by manufacturers in the United States and provided on their packaging, is the Noise Reduction Rating (NRR).[16] The NRR is an attenuation index that represents the overall average A-weighted noise reduction, in decibels, that an HPD will provide in an environment with a known C-weighted sound level. A similar descriptor has been standardized as the single-number rating (SNR) by the International Organization for Standardization (ISO).[17]

The NRR is calculated in a manner analogous to the OB approach, except that a pink noise spectrum (equal energy in each OB) is used instead of the actual spectrum. The computation includes a two-SD adjustment for percentage of population protected, as with the OB method, and an additional spectral safety factor of 3 dB.[4] The NRR computed for the HPD example in Table 1 is 20.7 dB.

The NRR is used to estimate wearer noise exposures by subtracting it from the C-weighted sound levels:

$$\text{Estimated exposure}[\text{dB}(A)] = \text{workplace noise level}[\text{dB}(C)] - \text{NRR}. \qquad (1)$$

The practice of subtracting the NRR from the C-weighted sound level may seem illogical; however, it is justified on theoretical and empirical grounds. Considerable accuracy is lost when the NRR is subtracted from A-weighted sound levels, in which case an additional safety factor of 7 dB must be included, as shown in Eq. (2). An alternative to the 7-dB correction has been provided by Berger[4]:

$$\text{Estimated exposure}[\text{dB(A)}] = \text{workplace noise level}[\text{dB(A)}] - (\text{NRR} - 7\text{dB}). \qquad (2)$$

7.3 Comparison of Octave Band and NRR Estimates

Although errors may arise in using the NRR to estimate results computed using the OB method, the NRR is usually of sufficient accuracy considering the *in*accuracies in the basic OB data from which either set of computations are made. For the example shown in Table 1, with exterior C- and A-weighted sound levels of 95.2 and 93.5 dB, respectively, the OB-computed A-weighted protected exposure level is 73.0 dB, as compared to the following NRR-computed values:

$$95.2\ \text{dB(C)} - 20.7 = 74.5\ \text{dB(A)}, \qquad (3)$$

or

$$93.5\ \text{dB(A)} - (20.7 - 7) = 93.5 - 13.8 = 79.8\ \text{dB}. \qquad (4)$$

Note the substantially larger discrepancy between the OB value and the estimate from Eq. (4) than that from Eq. (3). This is due to the conservative 7-dB adjustment required when only A-weighted sound levels are available.

The primary reason for examining the OB data is to better match the attenuation curve of the device to the noise spectrum. For example, both laboratory and workplace data show that earmuffs are a poor choice for noises with significant low-frequency energy (125–250 Hz), in which case a foam or premolded earplug is preferred, but are best when strong midfrequency energy is present (primarily around 1000 Hz).

7.4 Derating Laboratory Data

The NRR, or the laboratory data from which it is computed, must be reduced (derated) to provide attenuation values more realistic for the workplace. A derating of the NRR by 10 dB is the minimum average correction necessary to achieve this goal;[4] however, even larger deratings or a percentage (such as 50%) instead of constant-decibel approaches may be justified and may be more appropriate. Some authors have suggested that a percentage derating should be less for the easier to fit HPDs such as earmuffs than for the harder to fit devices such as earplugs.[13]

Whatever derating is selected, it must be applied equally to the OB calculations of attenuation; that is, if a percentage derating is used for NRRs, that same percentage should also be taken from the OB attenuation values before they are entered in row 4 of Table 1.

7.5 OSHA's Procedure for Derating Laboratory Data

The Occupational Safety and Health Administration (OSHA) specifies use of manufacturers' labeled data derived from laboratory measurements to assess adequacy of HPDs for the noise exposures in which they are worn.[1] Subsequent to publication of the Hearing Conservation Amendment, OSHA recognized that laboratory data required derating and recommended reducing published NRRs by 50% (e.g., an NRR of 24 dB would be reduced to 12 dB), but only for the purpose of evaluating the *relative efficacy* of HPDs and engineering noise controls.[18] The derated NRRs are not applicable for determining compliance with the hearing protection requirements of the Hearing Conservation Amendment.

8 EFFECTS OF HPDs ON AUDITORY PERCEPTION AND INTELLIGIBILITY

8.1 Speech Communications

Conventional hearing protectors cannot differentiate and selectively pass speech versus noise at a given frequency; thus the devices do not improve the speech-to-noise ratio, which is the most important factor in achieving reliable intelligibility. However, hearing protectors sometimes do afford intelligibility improvements in intense noise by lowering the total energy of both speech and noise incident on the ear. Overload distortion in the cochlea is thereby reduced, establishing more favorable conditions for discrimination.

Prediction of the effects of protectors on speech intelligibility in noise is a complex issue that depends on a host of factors, including the listener's hearing abilities, absolute speech and noise levels and speech-to-noise ratio, whether or not the talker is occluded, HPD attenuation, reverberation time of the environment, facial cues, and content/complexity of the message.

A distillation of the evidence concerning speech communication for *normal hearers* with HPDs suggests that conventional passive HPDs have little or no degrading effect on intelligibility in noise above about 80 dB(A) but cause considerable misunderstanding at lower levels, for which the use of protection is not typically needed anyway. At ambient noise levels greater than about 85 dB(A), several studies have reported slight intelligibility improvements with specific HPDs,[19] while others attempting to simulate on-the-job conditions have reported small intelligibility decrements, especially when the talker's ears are occluded.[20]

When wearing HPDs in noise, one tends to lower the voice level because the bone-conducted voice feedback inside the head is amplified by the presence of the protector, especially at low frequencies. In comparison, attenuation of the ambient noise occurs primarily through interruption of air conduction by the HPD. Thus, one's own voice is perceived as louder in relation to the noise than is actually true, causing a compensatory lowering of about 2–4 dB unless a conscious effort is made to "speak up."[4]

For individuals with high-frequency hearing losses, the effects of HPDs on communications are not clear-cut, although these persons are certainly at a disadvantage due to the fact that their already elevated thresholds for mid- to high-frequency speech sounds are further raised by the protector. As the amount of hearing loss increases, the chances of degrading communications with HPDs also increases. Though there is not consensus among studies, it does appear that hearing-impaired individuals often experience no improvement or even reduced communication abilities with HPDs in noise environments from 80 to 95 dB(A).

8.2 Auditory Warning Signals and Other Sounds

As with speech communication effects, the same HPD influences on signal-to-noise ratio and reduction of aural distortion apply to the detection and recognition of nonverbal signals such as warnings, target annunciators, and machinery sounds. Due to the high-frequency bias in attenuation of conventional HPDs, coupled with the typically elevated high-frequency thresholds of the neurally hearing-impaired, signals above about 2000 Hz are the most likely to be missed. However, warning signal parameters, such as frequency, intensity, and temporal profile, may be designed to help alleviate detection problems and increase perceived urgency. For instance, low-frequency (i.e., below 500 Hz) warning signals more readily diffract around barriers and are less attenuated by HPDs, thereby aiding detection.

Although aural cues may sound spectrally different under an HPD, the bulk of empirical studies indicates that signal detection will not be compromised by HPDs for normal-hearing individuals.[21] While the evidence is less extensive for hearing-impaired listeners, they can be expected to experience some detection and recognition difficulty, depending on the particular signal, ambient noise, and hearing protector worn. When an occluded/unoccluded detection disadvantage does occur, it is more common with earmuffs than earplugs, perhaps due to the muffs' stronger tendency to attenuate high-frequency signals and to pass more low-frequency noise, which effects an upward spread of masking.

8.3 Localization

Perceptual judgments of sound direction and distance may be influenced by HPDs, at least partially because some of the high-frequency binaural cues (above about 4000 Hz) that depend on the pinnae are lost or altered. Earmuffs and helmets, which completely obscure the pinnae, interfere most with localization in the vertical plane and also tend to cause contralateral (left-right) and ipsilateral (front-back) judgment errors. Earplugs generally induce fewer localization problems than muffs.

8.4 General Remarks

In most hazardous noise exposure situations, even though the quality of the auditory experience may change under HPDs, detection and intelligibility performance is not hampered for individuals of normal hearing. While hearing-impaired individuals may exhibit a reduced communication ability under HPDs in noise, use is not contraindicated; instead, it is essential that the remaining auditory sensitivity be preserved by proper protector use. All individuals may benefit from visually or tactually presented cues to augment auditory information, or in certain situations from the use of HPDs incorporating special communications features as described next.

9 HPDs WITH SPECIAL FEATURES

Special hearing protector designs have been developed to improve speech communication and auditory perception capabilities of the noise-exposed wearer. These devices may be categorized into passive (i.e., without electronics), active (i.e., incorporating electronics), and communications headsets.

9.1 Passive HPDs

Frequency-Sensitive HPDs Efforts to improve communications under earplugs have involved the use of apertures or channels through an earplug body, sometimes opening into an air-filled cavity encapsulated by the earplug walls. For example, a small (~0.5-mm) longitudinal channel through a custom-molded earplug causes an air leak that creates a low-pass filter characteristic. Attenuation is negligible below about 1000 Hz, increasing up to about 30 dB at 8000 Hz. Because most of the speech frequencies critical to intelligibility lie in the 1000–4000-Hz range, the potential communication benefits of the low-pass feature may be relatively small, especially in low-frequency noise, which causes upward masking into the critical speech band. Furthermore, the attenuation may be insufficient for many industrial noise environments.

Amplitude-Sensitive HPDs Because a conventional HPD provides a constant amount of attenuation that is independent of incident sound level, hearing ability is compromised during the quiet periods of intermittent sound exposures. Amplitude-sensitive HPDs address this problem by providing reduced attenuation at low sound levels, with increasing protection at high levels.

Passive HPDs utilize a nonlinear component, such as a sharp-edged orifice opening into a duct, to effect the change in attenuation. This technique, used with success in both earmuffs and earplugs, takes advantage of the fact that low-intensity sound waves predominantly exhibit laminar airflow and pass relatively unimpeded through the aperture, whereas high-intensity waves involve turbulent flow and are attenuated due to increasing acoustic resistance.[22]

A critical performance parameter is the transition sound level, normally 110–120 dB, above which insertion loss increases at a rate of about 1 dB for each 2–4-dB increase in sound level. Because this transition level is so high, passive amplitude-sensitive HPDs are best suited for outdoor impulsive blasts and gunfire exposures. At lower but still hazardous sound levels, most passive devices exhibit vented earplug behavior, affording very weak protection at frequencies less than 1000 Hz. One exception is an orifice-type earmuff that provides approximately 25 dB attenuation from 400 to 8000 Hz.[22]

Uniform Attenuation HPDs Because the attenuation of conventional HPDs increases with frequency, not only are sounds reduced in level, but the wearer's perception of spectral balance is distorted as well. This problem contributes to protector nonuse where, for instance, a machine operator's auditory feedback from a cutting tool is muffled or a musician's perception of timbre is compromised. To counter such problems, flat or uniform attenuation HPDs are designed to impose attenuation that is nearly constant from about 100 to 8000 Hz. One successful technique employs a Helmholtz resonator and a sound channel through a custom-molded earplug to provide a flat attenuation profile of about 15 dB.[23]

Properly fit uniform attenuation HPDs provide adequate protection and better hearing perception in low to moderate noise levels of about 90 dB(A) or less. Workers with high-frequency hearing losses and professional musicians may find them particularly beneficial.

9.2 Active HPDs

Amplitude-Sensitive Sound Transmission HPDs

Active sound transmission hearing protectors consist of modified conventional earmuffs or earplugs that incorporate microphone and limiting amplifier systems to transmit external sounds (often filtered to include the critical speech frequencies) to earphones mounted inside the earcups. Typically, the amplifier maintains a predetermined (in some cases user-adjustable) earphone level, often at about 85 dB(A), until the ambient noise is so intense that direct transmission through the earcup becomes the controlling factor. In comparison to both conventional and passive amplitude-sensitive earmuffs, sound transmission earmuffs are more expensive (upward of $100) but offer a viable alternative for use in intermittent noises, especially those with impulse-type (e.g., gunfire) or short-duration on-segments. However, in continuous, high-level noise, the distortion products of some systems may cause annoyance and compromise hearing ability.

Auditory perception and noise level under sound transmission earmuffs depends on numerous electronic design parameters. The microphones may be diotic, wherein a single microphone in one earcup feeds both earphones, or dichotic, in which each earcup has an independent microphone. The latter approach provides better localization performance.

Active Noise Reduction HPDs Recently integrated with conventional passive earmuffs, active noise reduction (ANR) electronics rely on the principle of destructive interference to cancel noise. A microphone senses the sound inside the earcup, which is then fed back through a phase compensation filter to a processing circuit and amplifier (Fig. 5). The resultant "antinoise" signal is presented through an earphone at equal amplitude to but 180° out of phase with the original noise, causing energy cancellation.

The ANR is most effective against repetitive or continuous noises that are relatively invariant in spectrum or level, which allow the system to stabilize and fine-tune the phase and amplitude parameters needed for cancellation. The ANR is limited to the reduction of low-frequency noises below about 1000 Hz, with maximum attenuation of 20–25 dB occurring below 300 Hz. Like conventional earmuffs, circumaural ANR devices are also susceptible to leakage under the earcup cushions such as caused by eyeglass temples, which can result in a reduction in active attenuation.[24]

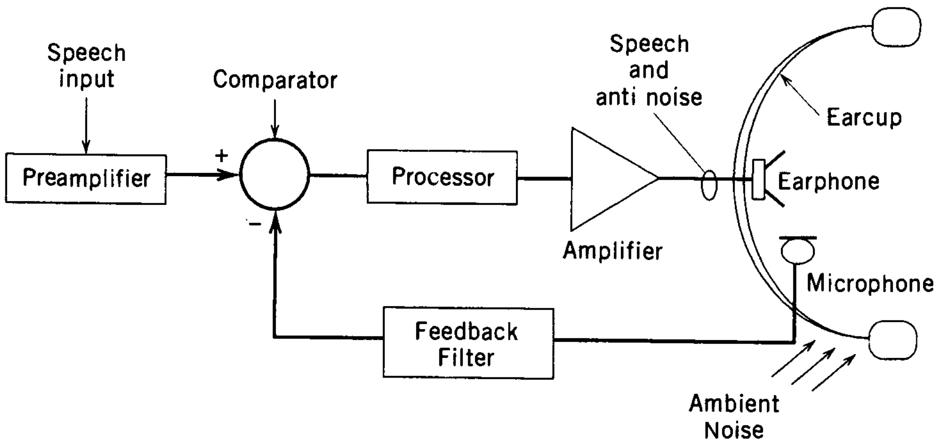


Fig. 5 Block diagram of an active noise reduction (ANR) communications headset.

The low-frequency effectiveness of ANR is fortuitous in that it compensates the typically high-frequency biased protection of a passive earmuff. If the noise environment consists of *only* low-frequency energy, open-back ANR headphones may offer communications and perhaps comfort advantages over circumaural ANR earmuffs. Consisting of lightweight supra-aural earphones that rest on the pinnae, these devices have the disadvantage of offering no passive protection if the ANR circuit malfunctions.

Attenuation tests for ANR hearing protectors are as yet not standardized. A combination procedure may be required, that is, REAT evaluations[8] to obtain the muff's passive attenuation and MIRE tests on human subjects[10] to obtain the active attenuation. The benefits of ANR devices come at a relatively high price of $150–$1000 per unit.

9.3 HPDs with Communications Features

For provision of communication or music signals at the ear, small loudspeakers have been integrated into HPDs. Headsets (including ANR examples) consist of earphones housed in earmuff earcups, which support a directional microphone (often noise canceling and/or voice activated) in front of the mouth. Small receivers can also be remotely located on a hard hat or behind the pinna and coupled to the ear via tubing through earplugs. An alternative is an earplug-like unit, called an ear microphone, which consists of a receiver button and a microphone that picks up the wearer's voice as a result of sound radiation from the bone-conduction-excited earcanal walls. Each of these approaches is available as one- or two-way systems using wireless (radio frequency or infrared) or wired technology.

It is important that communications earphones and receiver buttons be output limited so that amplified signal levels do not pose a hearing hazard. Furthermore, care must be exercised when selecting such devices for hearing protection purposes, since although some circumaural devices provide passive attenuation equivalent to a standard earmuff, certain ear microphones have been shown to provide less protection than comparable conventional earplugs.[25]

In cases where the attenuation afforded by a circumaural headset is inadequate, so that communications signals are noise masked, improvements can be realized by wearing an earplug under the headset. While the earplug will reduce the communications signal as well as the noise, the signal-to-noise ratio in the earcanal can be improved if the earphone provides sufficient distortion-free gain to compensate for the insertion loss of the earplug. Other enhancements may be provided in the system electronics, such as peak clipping and signal conditioning, to enhance the acoustic power of consonants, which are critical to word discrimination.

As diagrammed in Fig. 5, ANR has been integrated into communications headsets. The sound (speech and noise) signal from the earcup microphone is fed back to a comparator through a filter that reverses the phase. The comparator's difference output is the undesirable noise component that is out of phase with the noise inside the earcup. Since the low-frequency components of the (desired) speech signal are also canceled by the ANR, a speech preamplifier is used for compensation. For certain noise situations, ANR headsets result in an improvement in speech intelligibility over the use of passive devices alone.[24]

10 HPD STANDARDS, REGULATIONS, AND GOVERNMENT LISTINGS

10.1 United States

National standards in the United States specify experimental procedures for measuring HPD attenuation, either via REAT[8] or a supplemental physical method using an acoustical test fixture.[10] The U.S. Department of Commerce has established a National Voluntary Laboratory Accreditation Program (NVLAP) that can accredit laboratories for conformance with the aforementioned REAT standard.

10.2 International Organization for Standardization

The ISO provides methods corresponding to the two ANSI documents cited above: a REAT procedure[9] and an acoustical test fixture method[26] as well as a standard for using REAT data to compute effective noise levels when HPDs are worn.[17] Numerous additional ISO and European standards on methods of approval for earmuffs, earplugs, and other HPDs, and guidelines for selection and care of HPDs are expected to be promulgated in the late 1990s.

10.3 Environmental Protection Agency

The EPA promulgated a regulation in 1979[16] that specifies that any device or material capable of being worn on the head or in the earcanal that is sold wholly or in part on the basis of its ability to reduce sound entering the ear is to be evaluated according to the REAT portion of ANSI S3.19[27] (a currently out-of-date standard) and to be labeled with an NRR. All commercially available HPDs sold in the United States must be labeled in accordance with this regulation.

10.4 Occupational Safety and Health Administration

OSHA requires that HPDs reduce the effective sound level at the ear to a time-weighted average level of less than or equal to 90 dB(A), except in the case of employees demonstrating a standard threshold shift, in which case the level must be less than or equal to 85 dB(A).[1] The methods of computing noise reduction conform to those described in Sections 7.1, 7.2, and 7.5.

10.5 Other Agencies

The National Institute for Occupational Safety and Health publishes compendia, which are periodically updated, of hearing protector data and specifications that are obtained from the manufacturers of the devices.[28] The Mining Safety and Health Administration also publishes a compilation, albeit more abbreviated.[29]

REFERENCES

1. OSHA, "Occupational Noise Exposure; Hearing Conservation Amendment," Occupational Safety and Health Administration, *Fed. Regist.*, Vol. 48, No. 46, 1983, pp. 9738–9785.
2. L. H. Royster and J. D. Royster, "Hearing Protection Devices," in A. S. Feldman and C. T. Grimes (Eds.), *Hearing Conservation in Industry*, Williams and Wilkins, Baltimore, 1985, pp. 103–150.
3. E. H. Berger, "EARLog #17—Ear Infection and the Use of Hearing Protection," *J. Occup. Med.*, Vol. 27, No. 9, 1985, pp. 620–623.
4. E. H. Berger, "Hearing Protectors—Specifications, Fitting, Use, and Performance," in D. M. Lipscomb (Ed.), *Hearing Conservation in Industry, Schools, and the Military*, College-Hill Press, Boston, 1988, pp. 145–191.
5. J. G. Casali, "Comfort: The 'Other' Criterion for Hearing Protector Design and Selection," in *Proceedings of Hearing Conservation Conf.*, Off. Eng. Serv., Univ. Kentucky, Lexington, KY, 1992, pp. 47–53.
6. E. H. Berger, "Review and Tutorial—Methods of Measuring the Attenuation of Hearing Protection Devices," *J. Acoust. Soc. Am.*, Vol. 79, No. 6, 1986, pp. 1655–1687.
7. E. H. Berger, "Development of a Laboratory Procedure for Estimation of the Field Performance of Hearing Protectors," in *Proceedings of Hearing Conservation Conf.*, Off. Eng. Serv., Univ. Kentucky, Lexington, KY, 1992, pp. 41–45.
8. ANSI, "Method for the Measurement of the Real-Ear Attenuation of Hearing Protectors," Am. Natl. Stds. Inst., S12.6-1984, New York, 1984.
9. ISO, "Acoustics—Hearing Protectors—Part 1: Subjective Method for the Measurement of Sound Attenuation," ISO 4869-1 : 1990, Geneva, Switzerland, 1990.
10. ANSI, "Microphone-in-Real-Ear and Acoustic Test Fixture Methods for the Measurement of Insertion Loss of Circumaural Hearing Protection Devices," S12.42-1995, Am. Natl. Stds. Inst., New York, 1995.
11. J. G. Casali and S. T. Lam, "Effects of User Instructions on Earmuff/Earcap Sound Attenuation," *J. Sound Vib.*, Vol. 20, No. 5, 1986, pp. 22–28.
12. E. H. Berger, J. R. Franks, and F. Lindgren, "International Review of Field Studies of Hearing Protector Attenuation," in A. Axelsson, H. Borchgrevink, R. P. Hamernik, L. Hellstrom, D. Henderson, and R. J. Salvi (Eds.), *Scientific Basis for Noise-Induced Hearing Loss*, Thieme Medical, New York, 1996.
13. M. Y. Park and J. G. Casali, "A Controlled Investigation of In-Field Attenuation Performance of Selected Insert, Earmuff, and Canal Cap Hearing Protectors," *Hum. Factors*, Vol. 33, No. 6, 1991, pp. 693–714.
14. E. H. Berger, "Laboratory Attenuation of Earmuffs and Earplugs Both Singly and in Combination," *Am. Ind. Hyg. Assoc. J.*, Vol. 44, No. 5, 1983, pp. 321–329.
15. J. Starck and J. Pekkarinen, "Objective Measurements of Hearing Protector Attenuation for Weapon's Noise Exposure, in *Proceedings of Hearing Conservation Conf. 1992*, Off. Eng. Serv., Univ. Kentucky, Lexington, KY, 1992, pp. 89–92.
16. EPA, "Noise Labeling Requirements for Hearing Protectors," Environmental Protection Agency, *Fed. Regist.*, Vol. 44, No. 190, 40CFR Part 211, 1979, pp. 56130–56147.
17. ISO, "Acoustics—Hearing Protectors—Part 2: Estimation of Effective A-Weighted Sound Pressure Levels When Hearing Protectors are Worn," ISO 4869-2 : 1994, Geneva, Switzerland, 1994.
18. OSHA, "OSHA Instruction CPL 2-2.20A, change 2, March 1," in *Industrial Hygiene Technical Manual*, U.S. Government Printing Office, Washington, D.C., 1987, pp. VI-13–VI-20.
19. J. G. Casali and M. J. Horylev, "Speech Discrimination in Noise: The Influence of Hearing Protection," in *Proceedings of the Human Factors Society—31st Annual Meeting*, Human Factors Soc., New York, 1987, pp. 1246–1250.
20. H. Hormann, G. Lazarus-Mainka, M. Schubeius, and H. Lazarus, "The Effect of Noise and the Wearing of Ear Protectors on Verbal Communication," *Noise Control Eng. J.*, Vol. 23, No. 2, 1984, pp. 69–77.
21. P. Wilkins and A. M. Martin, "Hearing Protection and Warning Sounds in Industry—A Review," *Appl. Acoust.*, Vol. 21, No. 4, 1987, pp. 267–293.
22. C. H. Allen and E. H. Berger, "Development of a Unique Passive Hearing Protector with Level-Dependent and Flat Attenuation Characteristics," *Noise Control Eng. J.*, Vol. 34, No. 3, 1990, pp. 97–105.
23. M. Killion, E. DeVilbiss, and J. Stewart, "An Earplug with Uniform 15-dB Attenuation," *Hearing J.*, Vol. 41, No. 5, 1988, pp. 14–17.

24. C. W. Nixon, R. L. McKinley, and J. W. Steuver, "Performance of Active Noise Reduction Headsets," in A. L. Dancer, D. Henderson, R. J. Salvi, and R. P. Hamernik (Eds.), *Noise-Induced Hearing Loss*, Mosby-Year Book, St. Louis, 1992, pp. 389–400.

25. D. Mauney and J. G. Casali, "Preliminary Noise Attenuation Assessment of AICOMM's AIMic: Current Fitting Device Performance and Improved Ear Couplers," Virginia Tech. Ind. Eng. and Operations Res., Rept. 9001, Blacksburg, VA, 1990.

26. ISO, "Acoustics—Hearing Protectors—Part 3: Simplified Method for the Measurement of Insertion Loss of Ear-Muff Type Protectors for Quality Inspection Purposes," ISO/TR 4869-3 : 1989, Geneva, Switzerland, 1989.

27. ANSI, "Method for the Measurement of Real-Ear Protection of Hearing Protectors and Physical Attenuation of Earmuffs," S3.19-1974 (ASA STD 1-1975), Am. Natl. Stds. Inst., New York, 1974.

28. J. R. Franks, C. L. Themann, and C. Sherris, "The NIOSH Compendium of Hearing Protection Devices," U.S. Dept. of HHS, Pub. No. 95-105, Cincinnati, OH, 1994.

29. MSHA, "Hearing Protector Factor List," Mining Safety and Health Administration, Pittsburgh, PA., 1996.

82

DETERMINATION OF NOISE EMISSION OF SOUND SOURCES

GEORGE C. MALING, JR. AND WILLIAM W. LANG

1 INTRODUCTION

To develop measures for reducing the noise of a source, it is necessary to have a quantitative measure of its noise emission. In the case of moving sources such as aircraft and motor vehicles, it is usual to describe the noise emission of the source in terms of the sound pressure level measured at specified microphone positions and under carefully prescribed environmental conditions. For stationary sources, especially sources that can be placed in a well-defined test environment, the sound power level of the source is a more commonly used measure of noise emission. The sound power of the source must, in some cases, be supplemented by a measure of the directivity of the source, the variation of sound pressure level with position around the source.

Standardized methods for determining the sound power levels of stationary sources are available, and are described in this chapter. Sound power may be determined directly from measuring the sound intensity on a measurement surface that surrounds the source (see Chapter 156) or it may be determined by measurement of the sound pressure level produced by the source under controlled environmental conditions. The methods in this chapter are restricted to the latter and to methods that have been standardized both nationally and internationally.

2 NOISE EMISSION, NOISE IMMISSION, AND SOURCE DIRECTIVITY

2.1 Noise Emission and Noise Immission

The sound pressure level, expressed in decibels, is the physical quantity nearly always used to describe the effects of noise on people because the ear is responsive to the pressure of a sound wave. The sound pressure level can be easily measured by a sound level meter and is used as the descriptor for *noise immission*, the sound pressure level at a specified position in the vicinity of a receiver: a microphone or a person. The sound pressure level varies with distance from a sound source and is highly dependent on the acoustical environment in which the source is placed. For these reasons, the sound pressure level, a good measure of the noise immission of a source, is not usually a good measure of the noise emission of that source unless the measurement positions in the vicinity of the source and the acoustical characteristics of the environment are carefully specified.

The *noise emission* of a source can be described in terms of the sound power level emitted by the source and its directivity.* The sound power level is a measure of the total sound power or sound energy per unit time radiated by the source and is usually expressed in decibels using a reference sound power of one picowatt (10^{-12}

---

*This chapter is an edited version of Chapter 4, "Determination of Sound Power Levels and Directivity of Noise Sources," by George C. Maling, Jr., William W. Lang, and Leo L. Beranek, which appeared in *Noise and Vibration Control Engineering*, Leo L. Beranek and Istvan L. Ver, Editors. Copyright © 1992 John Wiley & Sons, Inc.

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker
Copyright © 1997 John Wiley & Sons, Inc.

W). The sound power level is a function of frequency and is often specified in octave or one-third-octave bands.[1] Since sound power level and sound pressure level may both be expressed in decibels (with different reference quantities), the potential for confusion between these two levels is very high. To make a clear distinction between sound pressure levels and sound power levels, the former is always expressed in decibels, while the latter may be expressed in *bels* (1 bel = 10 dB).

2.2 Source Directivity

The directivity of a source is a measure of the variation of sound radiation with direction around a source. In principle, the directivity of a source can be described by the sound pressure level radiated as a function of the angular position around the source in a specified acoustical environment such as a free field. The directivity is also a function of frequency. Consequently, a complete description of source directivity is complex, and in practice the directivity index is usually determined in one or more specified directions and for a frequency-weighted spectrum—for example, with A-weighting.

The directivity of a source and the sound power level of that source are both influenced by the presence of nearby reflecting surfaces; in most practical situations, the influence of reflecting surfaces on the sound power level is small except in unusual situations. For example, the sound power level of a discrete-frequency source in a highly reflective environment such as a reverberation room is strongly dependent on source position when the source excites relatively few normal modes in the room. In many cases of practical interest, the directivity of the source can be described by a single number, and the descriptor is the sound pressure level at a given distance from the source. Outdoors, the relationship between sound pressure level and sound power level depends on distance, the acoustical characteristics of the terrain, and meteorological conditions such as temperature, wind speed, and atmospheric turbulence. Indoors, the size and shape of the room in which the source is located, as well as the acoustical characteristics of the room surfaces, greatly influence the difference between the sound power level of the source and sound pressure level at a given receiver position in the vicinity of the source.

3 A-WEIGHTED SOUND POWER LEVEL

When A-weighting is applied to the sound power level spectrum of a source, the result is the *noise power emission level* (NPEL) of the source, the A-weighted sound power level in bels. This provides a single-number rating for source noise emission. The A-weighted sound power level in bels is defined as

$$L_{WA} = \log \frac{W_A}{W_0}, \tag{1}$$

where W_A is the A-weighted sound power of the source and W_0 is the reference sound power, standardized as 10^{-12} W. Use of the bel as the unit of sound power level avoids confusion between the noise emission of a sound source and noise immission, the sound pressure level at a given receiver position in the sound field. This distinction is particularly important when dealing with the general public and with persons who are not familiar with the technical details of acoustics. Engineers who have a clear understanding of the difference between sound power and sound pressure frequently express the sound power level of a source in decibels as

$$L_W = 10 \log \frac{W}{W_0}, \tag{2}$$

where W is the sound power of the source, unweighted or in a specified frequency band.

4 SOUND POWER, SOUND INTENSITY, AND SOUND PRESSURE

The sound power radiated by a source can be calculated from the sound intensity on a closed surface that completely surrounds the source as follows:

$$W = \int_S \mathbf{I} \cdot d\mathbf{S}, \tag{3}$$

where W is the sound power in watts, $\mathbf{I}$ the sound intensity in watts per square centimetre, $d\mathbf{S}$ the element of surface area in square metres, and S the surface that surrounds the source.

The surface integral of Eq. (3) is the integral of the component of sound intensity normal to an element of surface area, $d\mathbf{S}$. The integral may be carried out over any surface that surrounds the source; spherical, hemispherical, or rectangular surfaces are most commonly used. It is important to note that the intensity *normal* to the measurement surface is integrated. In general, this is not the maximum intensity at each point.

In practice, the integral of Eq. (3) is replaced by a summation; the surface S is divided into N segments, the ith segment having an area S_i. Equation (1) is then approximated by the summation

$$W = \sum_{i=1}^{N} I_i S_i. \tag{4}$$

Here, I_i is the sound intensity normal to the ith surface segment and S_i is the area of the ith surface segment.

The sound power level can then be obtained from Eq. (2).

In many cases of practical interest, each of the subareas, S_i, has the same area. Equations (4) and (2) can then be used to express the sound power level L_W as

$$L_W = L_I + 10 \log \frac{S}{S_0}, \tag{5}$$

where I is the average sound intensity on the measurement surface, $L_I = 10 \log(I/I_0)$, where I_0 is the reference sound intensity (10^{-12} W/m^2), S is the area of the measurement surface in square metres, and $S_0 = 1$ m^2.

Note that when the average sound intensity is computed, intensity in the outward direction from the surface is taken as positive and intensity in the inward direction is taken as negative.

Equation (5) may be used to compute sound intensity except when the source is highly directive, and unequal area segments of high (or low) sound intensity must be identified; in this case, Eqs. (4) and (2) should be used. A *field indicator* may be used to determine how many measuring points on the measurement surface are required to achieve a given degree of accuracy in the determination of sound power via sound intensity. See Chapter 156.

Several national and international standards use an approximate relationship between sound pressure and sound intensity to determine the sound power level of the source. In the far field of a source radiating into a free field or a free field over a reflecting plane, the sound intensity may be approximated by

$$I = \frac{p_{\text{rms}}^2}{\rho c}, \tag{6}$$

where p_{rms}^2 is the mean-square sound pressure at the measuring point in newtons per square metre, ρ the density of air in kilograms per cubic metre, and c the speed of sound in metres per second.

Since the mean-square sound pressure is always positive, the assumption is made that the sound intensity approximated by Eq. (6) is always in the direction of the outward normal to the measurement surface and that it is the maximum sound intensity at the point. The sound pressure level L_p is defined as $10 \log(p_{\text{rms}}^2/p_0^2)$, where p_0 is the reference sound pressure (2×10^{-5} N/m^2).

Equations (5) and (6) may be combined to yield the sound power level in decibels relative to 10^{-12} W:

$$L_W = L_p + 10 \log \frac{S}{S_0} - 10 \log \frac{\rho c}{400}, \tag{7}$$

where L_p is the level of the average mean-square sound pressure on the measurement surface in decibels relative to 2×10^{-5} N/m^2.

In practice, the last term in Eq. (7) is usually small and can be neglected for measurements of engineering grade. For example, under normal conditions, the density of air is about 1.18 kg/m^3, and the speed of sound is about 344 m/s. Under these conditions, the last term in Eq. (7) is 0.06 dB. The term is retained in standards that require precision grade measurements since air density and temperature may differ from normal conditions. For sources that are highly directive, Eqs. (2), (4), and (6) can easily be combined to determine the sound power level in terms of sound pressure level using area segments that have unequal values.

The principal advantages of using sound intensity directly to determine sound power are due to the directional nature of sound intensity. See Section 5.5. When the approximation is made that the mean-square pressure at a point is proportional to sound intensity, these advantages are lost because the mean-square pressure is always positive; background noise levels, interfering sources, and room reflections generally add to the measured mean-square pressure. This means that special test environments are required to determine sound power from sound pressure. Good approximations to a free-field environment can be achieved in properly designed anechoic or hemi-anechoic rooms or outdoors.

The sound power of a source can also be computed from sound pressure measurements in an enclosure in which the sound field is approximately diffuse, usually achieved by placing the source under test in a reverberation room. In this case, the sound energy density in the room may be assumed to be constant; it is directly related to the mean-square sound pressure and, therefore, to the sound power radiated by the source. When a source is operated in a reverberation room, the sound pressure level in the room builds up until the energy absorbed by the room surfaces or other sound absorbers is equal to the energy produced by the source under test. The sound power produced by the source is determined by measuring the space-averaged mean-square pressure in the reverberant field and calculating the sound power level either directly (*direct method*) or by comparing the mean-square pressure with the mean-square pressure produced by a source of known sound power output, a reference sound source (*comparison method*).

5 STANDARDS FOR THE DETERMINATION OF SOUND POWER LEVEL

There are many published standards[2-13] that involve determination of the noise emission of sources in terms of sound power level. In this section, the various measurement environments are described, the basic equations used to determine sound power are given, and examples of microphone arrays are shown. Individual standards should be consulted for detailed measurement requirements.

5.1 Reverberation Room Methods

As discussed above, a reverberant environment may be used to determine the sound power level of a noise source from measurements of the space-averaged sound pressure level in a diffuse field. If the source emits broadband noise, ANSI S12.31[4] or ISO 3741[5] may be used. The basic equation used to determine the sound power level using the direct method is

$$L_W = L_p - 10\log\frac{T}{T_0} + 10\log\frac{V}{V_0} + 10\log\left(1 + \frac{S\lambda}{8V}\right) - 10\log\frac{B}{1000} - 14, \tag{8}$$

where L_p is the space-averaged sound pressure level of the equipment being evaluated in decibels relative to $2 \times 10^{-5} N/\mathrm{m}^2$, T the reverberation time of the room for the frequency band of interest in seconds (T_0 = 1 s), V the room volume in cubic metres ($V_0 = 1\ \mathrm{m}^3$), S the area of all boundary surfaces of the room in square metres, λ the wavelength corresponding to the center frequency in the octave or one-third-octave band being measured, and B the barometric pressure in millibars (1000 mbars = reference barometric pressure).

The sound power level of the source, L_{We}, determined using the comparison method is determined by

$$L_{We} = L_{Wr} + (L_{pe} - L_{pr}), \tag{9}$$

where L_{pe} is the space-averaged sound pressure level of the equipment being evaluated in decibels relative to $2 \times 10^{-5}\ \mathrm{N/m^2}$, L_{Wr} the sound power level of the (calibrated) reference sound source in decibels relative to 10^{-12} W, and L_{pr} the space-averaged sound pressure level of the reference sound source in decibels relative to $2 \times 10^{-5}\ \mathrm{N/m^2}$.

If the source under test radiates discrete-frequency components, special requirements are placed on the reverberation room to ensure that the accuracy of the spatial sampling and the influence of the room on the source sound power level conform to specified requirements. The procedures to be followed are described in ANSI S12.32 and ISO 3742. The standard ISO 3742 is being revised and combined with ISO 3741.

5.2 Methods in Special Test Rooms

The requirements for a special test room are described in ANSI S12.33.[8] This standard is useful for the determination of noise emission of small sources such as home appliances. The methods of ANSI 12.33 have been updated in ISO 3743,[9,10] and a comparison method that can be used in a diffuse-field environment has been added.

5.3 Engineering Methods in a Hemi-Anechoic Room

Engineering methods for determination of sound power under essentially free field conditions over a reflecting plane are described in ANSI S1.34[11] and ISO 3744.[12] The latter has recently been updated, and the requirements in the two documents are not identical. The measurements can be made in a laboratory-quality hemi-anechoic room or in a specially designed test room. A small (up to 2 dB) correction, K_1, may be subtracted from the right-hand side of Eq. (7) to account for environmental conditions. Detailed methods for determining the magnitude of the correction are given in the standards. The correction can be determined in several ways:

1. By comparing the calibrated sound power level of a reference source, L_{Wr}, with the measured sound power level of the same source in the room, L_W. The environmental correction is $K_1 = L_W - L_{Wr}$.
2. By measuring the reverberation time in the room and determining the total Sabine absorption A according to $A = 0.16(V/T)$, where T is the reverberation time in the room in seconds and V is the volume of the test room in cubic metres. The environmental correction K_1 is determined using Fig. 1. In the figure, S is the area of the measurement surface in square metres.
3. In ANSI S12.34, the environmental correction may also be determined using a two-surface method described in the standard.

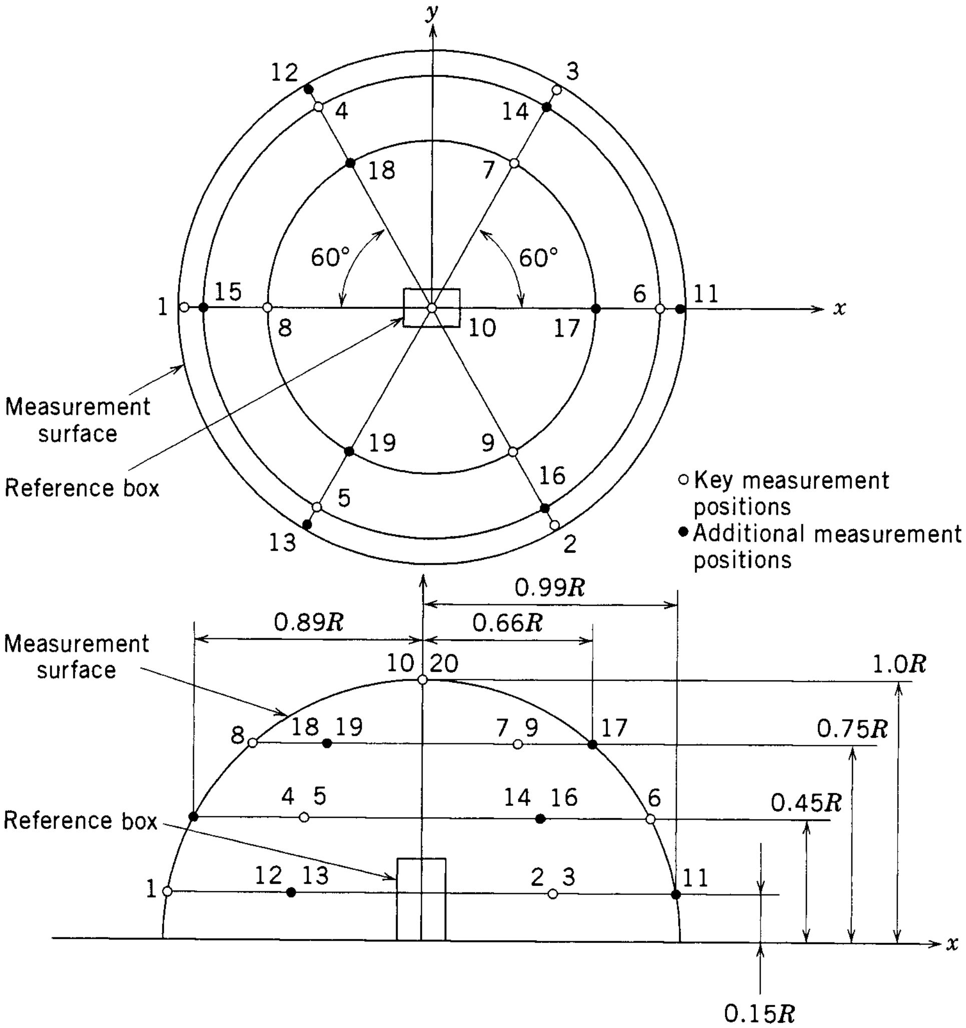

y
12
3
4
14
18
7
60°
60°
1
15
8
10
17
6
11
x
Measurement surface
Reference box
19
9
16
5
13
2
○Key measurement positions
●Additional measurement positions
0.99R
0.89R
0.66R
Measurement surface
10
20
1.0R
8
18 19
7 9
17
0.75R
4 5
14 16
6
0.45R
Reference box
1
12 13
2 3
11
x
0.15R

Fig. 1 Microphone array for a hemispherical surface according to ISO 3744. Key microphone positions are numbered 1–10. Additional microphone positions are numbered 11–20 and may be used to improve the accuracy of the measurements. (Reprinted from ANSI S12.34 by permission of the Acoustical Society of America.)

The source sound power level is determined from the average sound pressure level on the measurement surface by using Eq. (7), with the last term set equal to zero, and the environmental correction subtracted from the right-hand side of the equation:

$$L_W = L_p + 10 \log \frac{S}{S_0} - K_1. \qquad (10)$$

Two of the allowed measurement surfaces are the hemispherical surface and the rectangular surface; these surfaces together with the specified microphone positions are given in Figures 2 and 3, respectively. For large machines, additional microphones on the measurement surface must be used; these positions are specified in the standards. Corrections for background noise levels must also be made as specified in the standards.

5.4 Precision Methods in Anechoic or Hemi-Anechoic Rooms

Precision methods for determination of sound power are given in ANSI S12.35[13] and ISO 3745.[14] Equation (7) is used to determine the sound power level from the sound pressure level, and an equation is given for highly directive sources that require the selection of area segments having unequal areas. Measurement surfaces, microphone arrays, correction for background noise levels, and other requirements are given in the standards.

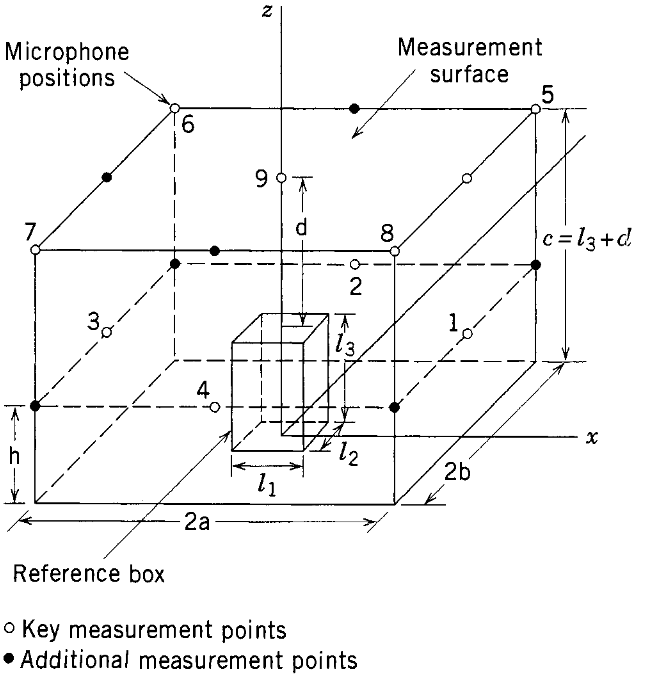


Fig. 2 Microphone array for a rectangular measurement surface according to ISO 3744. The distance d from the rectangular parallelepiped that surrounds the source to the measurement surface is often taken as 1 m. The height h given in the figure is usually $0.5(l_3 + d)$ but may be the shaft height for rotating machines, the center of the machine, or 1.2 m. The nine key microphone positions are usually adequate, but the additional microphone positions shown should be used if the major source dimension exceeds twice the distance d. (Reprinted from ANSI S12.34 by permission of the Acoustical Society of America.)

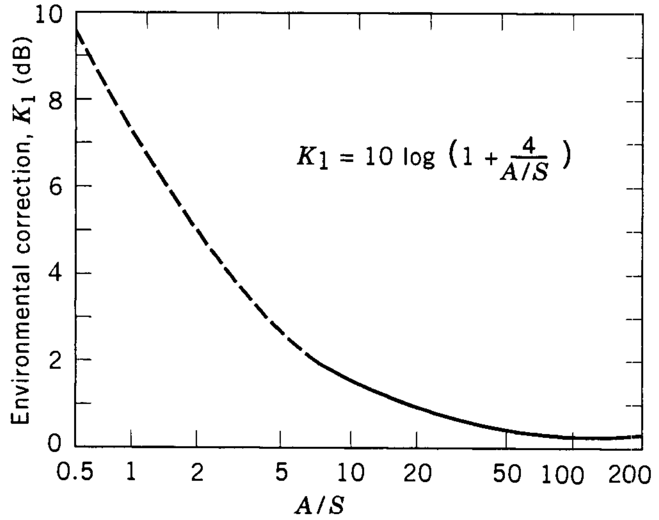


Fig. 3 Environmental correction K_1 in decibels. (Reprinted from ANSI S12.36 by permission of the Acoustical Society of America.)

5.5 Sound Intensity Methods in Ordinary Rooms

If an instrument for the measurement of sound intensity is available, it is convenient to determine sound power directly by using Eqs. (4) and (2). Determination of sound power directly from measurement of sound intensity has several advantages. The sound power of a source can often be determined in an ordinary room; special test environments may not be required. Sources that are outside the closed surface defined in Eq. (3) do not contribute to the surface integral, and, in principle, only the sound power of the source itself is determined. Also, background noise and reflections from room surfaces do not contribute to the surface integral. This means that the sound power of a source may, in principle, be determined in the presence of high background noise levels, in the presence of interfering sources, and in reflective acoustical environments such as ordinary rooms. In practice, these effects are not totally eliminated since the surface integral is approximated by a sum over finite segments of surface area. If the equipment being tested is sound absorptive, or if there is sound absorptive material inside the measurement surface, background noise, reflections, and sources outside the measurement surface may contribute to the surface integral in Eq. (3). Detailed methods for the determination of sound power from sound intensity are given in Chapter 156.

6 DETERMINATION OF SOURCE DIRECTIVITY

In principle, the directivity of a sound source can be specified by the sound pressure level, weighted or in a given frequency band, at a fixed distance and in different directions from the source. The data then constitute the directivity pattern of the source. In practice, the source directivity is usually specified in one or more directions in terms of the directivity factor Q. In a specified direction from the source, the directivity factor Q is calculated as

$$Q = \frac{p^2}{p_s^2}, \tag{11}$$

where p^2 is the mean-square pressure produced by a source of sound power W at a given distance from the source and in a specified direction and p_s^2 the mean-square pressure at the same distance from a nondirectional source that radiates the same sound power W. The directivity index is 10 log Q. The calculation of directivity factor and directivity index for several environmental situations is given elsewhere.[14]

REFERENCES

1. C. M. Harris (Ed.) "Measurement of Sound Power," in *Handbook of Acoustical Measurements and Noise Control*, McGraw-Hill, New York, 1991, Chap. 13.
2. ANSI S12.30-1990, "American National Standard Guidelines for the Use of Sound Power Standards and for the Preparation of Noise Test Codes," Acoustical Society of America, New York.
3. ISO 3740:1980, "Acoustics—Determination of Sound Power Levels of Noise Sources—Basic Standards and for the Preparation of Noise Test Codes," International Standardization Organization, Geneva, Switzerland.
4. ANSI S12.31-1990, "Precision Methods for the Determination of Sound Power Levels of Broad-band Noise Sources in a Reverberant Room," Acoustical Society of America, New York.
5. ISO 3741:1988, "Acoustics—Determination of Sound Power Levels of Noise Sources—Broad-band Sources in Reverberation Rooms," International Standardization Organization, Geneva, Switzerland.
6. ANSI S12.32-1990, "Precision Methods for the Determination of Sound Power Levels of Discrete-Frequency and Narrow-Band Noise Sources in Reverberation Rooms," Acoustical Society of America, New York,
7. ISO 3742:1988, "Acoustics—Determination of Sound Power Levels of Noise Sources—Discrete-Frequency and Narrow Band Sources in Reverberation Rooms," International Standardization Organization, Geneva, Switzerland.
8. ANSI S12.33-1990, "American National Standard Engineering Methods for the Determination of Sound Power Levels of Noise Sources in a Special Reverberation Test Room," Acoustical Society of America, New York,
9. ISO 3743-1:1994, "Acoustics—Determination of Sound Power Levels of Noise Sources Using Sound Pressure—Engineering Methods for Small, Movable Sources in Reverberant Fields—Part 1: Comparison Method in Hard-Walled Test Rooms," International Standardization Organization, Geneva, Switzerland.
10. ISO 3743-2:1994, "Acoustics—Determination of Sound Power Levels of Noise Sources Using Sound Pressure—Engineering Methods for Small, Movable Sources in Reverberant Fields—Part 2. Methods for Special Reverberation Test Rooms," International Standardization Organization, Geneva, Switzerland.
11. ANSI S12.34-1988 (R1990), "Engineering Methods for the Determination of the Sound Power Levels of Noise Sources for Essentially Free-field Conditions over a Reflecting Plane," Acoustical Society of America, New York,
12. ISO 3744:1994, "Acoustics—Determination of Sound Power Levels of Noise Sources—Free Field Conditions over a Reflecting Plane," International Standardization Organization, Geneva, Switzerland.
13. ANSI S12.35-1990, "American National Standard Precision Methods for the Determination of Sound Power Levels of Noise Sources in Anechoic and Hemi-anechoic Rooms," Acoustical Society of America, New York,
14. ISO 3745:1977, "Acoustics—Determination of Sound Power Levels of Noise Sources—Precision Methods for Anechoic and Semi-anechoic Rooms," International Standardization Organization, Geneva, Switzerland.
15. L. L. Beranek and I. Ver (Eds.), "Determination of Sound Power Levels and Directivity of Noise Sources," in *Noise and Vibration Control Engineering*, Wiley, New York, 1992, Chap. 4.

83

THE GENERATION OF NOISE IN MACHINERY, ITS CONTROL, AND THE IDENTIFICATION OF NOISE SOURCES

MALCOLM J. CROCKER

1 INTRODUCTION

Machines are used for a variety of purposes. The noise of appliances is often simply just annoying. The noise of machines in industry, however, is frequently intense enough to cause hearing damage. Although each machinery noise problem is somewhat different, a systematic approach and use of several well-known methods often produce sufficient noise reduction and acceptable acoustical conditions.[1] This chapter begins with a discussion of the *source–path–receiver* model, continues with a description of the most useful passive noise control approaches, and concludes with a discussion of the sources of noise in several important classes of machine and methods of identifying noise sources.

2 SYSTEMATIC APPROACH TO NOISE PROBLEMS

Noise control should always be attempted at the design stage wherever possible because there are more low-cost options and possibilities then to make individual machines or installations quieter.[2,3] After machines are built or installations completed, noise control approaches can still be achieved through various modifications and add-on treatments, but these are frequently more difficult and expensive to implement.

Noise problems can be described using the simple *source–path–receiver* model shown in Fig. 1.[4] The *sources* are of two main types: (1) airborne sound sources caused by gas fluctuations (as in the fluctuating release of gas from an engine exhaust) or (2) structure-borne machinery vibration sources that in turn create sound fields (for example, engine surface vibrations). Moreover, these sound pressure and vibration sources are of two types: (1) steady state and (2) impulsive. Both steady-state and impulsive vibrations (caused by impacting parts) are commonly encountered in machines. The *paths* may also be airborne or structure-borne in nature.

Source modifications are the best practice, but are sometimes difficult to accomplish. Often changes in the *path* or at the *receiver* may be the only real options available. The model shown in Fig. 1 is very simple. In reality there will be many sources and paths. The dominant source should be treated first, then the secondary one, and so on. The same procedure can also be applied to the paths. Finally, when all other possibilities are exhausted, the receiver can be treated. If, as in most noise problems, the receiver is the human ear, earplugs or earmuffs or even complete personnel enclosures can be used.

Measurements, calculations, and experience all play a part in determining the dominant noise sources and paths. The dominant sources (and paths) can sometimes be determined from careful experiments. In some cases, parts of a machine can be turned off or disconnected to help identify sources. In other cases, parts of a machine can be enclosed, and then sequential exposure

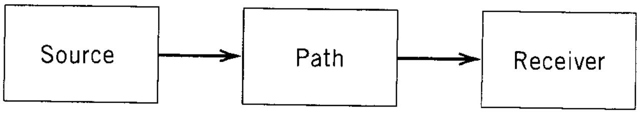


Fig. 1 Source–path–receiver model for noise problems.

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker
Copyright © 1997 John Wiley & Sons, Inc.

of machine parts can be used to identify major sources. Frequency analysis of machines can also be used as a guide to the causes of noise, as with the case of the firing frequency in engines, the pumping frequency of pumps and compressors, and the blade-passing frequency of fans. More sophisticated methods are also available involving the use of coherence, cepstrum, and intensity methods. Methods for determining the sources of noise in machinery are discussed in Section 6 of this chapter.

3 NOISE REDUCTION TECHNIQUES

A study of the literature reveals many successful well-documented methods used to reduce the noise of machines. These can be classified using the source–path–receiver model. Some of the most useful approaches can generally be used only at the source or in the path. Others, such as *enclosure*, can be adapted for use at any location. For instance, a small enclosure can be built inside a machine around a gear or bearing, or a larger enclosure or room can be built around a complete machine. Finally, an enclosure or personnel booth can be built for the use of a machine operator. Table 1 summarizes a large number of approaches that have been found useful in practice.

4 MAIN PASSIVE NOISE CONTROL APPROACHES

In Section 4 the main passive noise control approaches are briefly summarized. These include the use of (1) vibration isolators, (2) acoustical absorbing material, (3) enclosures, (4) barriers, and (5) vibration damping material.

4.1 Use of Vibration Isolators

The design of vibration isolators has been discussed frequently in the literature. Vibration isolators are used in two main situations: (1) where a machine source is producing vibration that it is desired to prevent flowing to supporting structures and (2) where a delicate piece of equipment (such as an electronics package or precision grinder) must be protected from vibration in the structure. It is the first case that will receive attention here. Primary emphasis is placed on reducing the force trans-

TABLE 1 Various Passive Noise Control Approaches That May Be Considered for Source, Path, or Receiver

Source	Choose quietest machine source available Reduce force amplitudes Apply forces more slowly Use softer materials for impacting surfaces Balance moving parts Use better lubrication Improve bearing alignment Use dynamic absorbers Change natural frequencies of machine elements Increase damping of machine elements Isolate machine panels from forces Put holes in radiating surfaces Stagger time of machine operations in a plant
Path	Install vibration isolators Use barriers Install enclosures Use absorbing materials Install reactive or dissipative mufflers Use vibration breaks in ductwork Mismatch impedances of materials Use lined ducts and plenum chambers Use flexible ductwork Use damping materials
Receiver	Provide earplugs or earmuffs for personnel Construct personnel enclosures Rotate personnel to reduce exposure time Locate personnel remotely from sources

mitted from the machine source to the supporting structure, but a secondary consideration is to reduce the vibration of the machine source itself.

It is often found that machines are attached to metal decks, grills, and sometimes lightweight wood or concrete floors. The machine on its own is usually incapable of radiating much noise (particularly at low frequency). The supporting decks, grills, and floors, however, tend to act like sounding boards, just as in musical instruments, and amplify the machine noise. Properly designed vibration isolators can overcome this noise problem.

Vibration isolators are of three main types: (1) spring, (2) elastomeric, and (3) pneumatic. Spring isolators are durable but have little damping. Elastomeric isolators are less durable and are subject to degradation due to corrosive environments. They have higher damping and are less expensive. Pneumatic isolators are used where very low frequency excitation is present.

Often the exciting forces are caused by rotational out-of-balance forces in machines or machine elements or are caused by magnetic or friction and other forces. Usually, these forces are simple harmonic in character. Such forces occur in electric motors, internal combustion engines, bearings, gears, and fans. Sometimes, however, the exciting forces may be impulsive in nature (e.g., in the case of punch presses, stamping operations, guillotines, tumblers, and any machines where impacts occur). The design of vibration isolators for a machine under the excitation of a simple harmonic force is considered here.

Theory of Vibration Isolation A machine may be considered, for simplicity, to be represented by a rigid mass m. If the machine is attached directly to a large rigid massive floor, as shown in Fig. 2, then all the periodic force $F_m(t)$ applied to the mass is directly transmitted to the floor. We will assume that the force on the mass is vertical and $F_m(t) = F_m \sin 2\pi f t$, where F_m is the amplitude of this force, and f is its frequency (Hz).

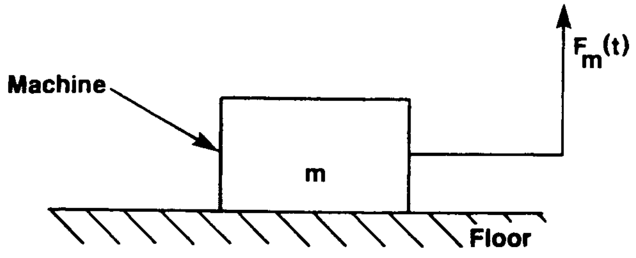


Fig. 2 Rigid machine of mass m attached to a rigid massive floor.

If a vibration isolator is now placed between the machine and the floor, we can model this system with the well-known single-degree-of-freedom system shown in Fig. 3.

Suppose, for the moment, that the periodic force is stopped; then both the applied force $F_m(t)$ and the transmitted force $F_t(t)$ will be zero. If the mass m is displaced from its equilibrium position and released, then it will vibrate with a *natural frequency* of vibration f_n given by

$$f_n = \left(\frac{1}{2\pi}\right)\sqrt{\frac{K}{m}} \quad \text{Hz}, \tag{1}$$

where, in SI units, K is the stiffness (N/m) and m is the mass (kg). Alternatively in English units, K is the stiffness [lb (force)/in.] and m is the mass [lb (mass)]. (Note: lb (force) = g lb (mass), where g = acceleration of gravity = 386.4 in./s^2.)

If the exciting force $F_m(t)$ is now resumed, then a force $F_t(t)$ will be transmitted to the rigid floor. This force $F_t(t)$ will be out of phase with $F_m(t)$, but it is sim-

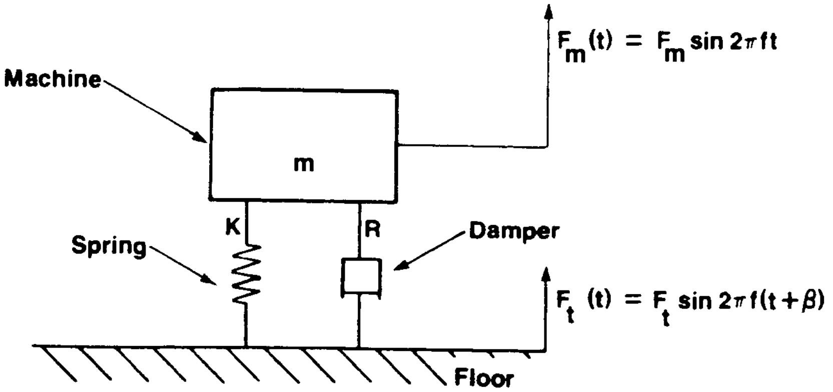

$F_m(t) = F_m \sin 2\pi ft$
Machine
m
Spring
K
R
Damper
$F_t(t) = F_t \sin 2\pi f(t+\beta)$
Floor

Fig. 3 Rigid machine of mass m separated from rigid massive floor by vibration isolator of stiffness K and damping R.

ple to show that the ratio of the amplitudes of the forces is given by

$$T_F = \frac{F_t}{F_m} = \sqrt{\left\{\frac{1 + 4(f/f_n)^2(R/R_c)^2}{[1 - (f/f_n)^2]^2 + 4(f/f_n)^2(R/R_c)^2}\right\}}, \tag{2}$$

where f is the frequency of the exciting force (Hz), f_n is the natural frequency of vibration of the mass m on the spring (Hz), R is the coefficient of damping, and R_c is the coefficient of critical damping, [$= 2\sqrt{mK}$]. The ratio F_t/F_m is known as the force transmissibility T_F.

The vibration amplitude A of the machine mass m is given by

$$\frac{A}{F_m/K} = \frac{1}{\{[1 - (f/f_n)^2]^2 + 4(f/f_n)^2(R/R_c)^2\}^{1/2}}. \tag{3}$$

The ratio $A/(F_m/K)$ is known as the *dynamic magnification factor* (DMF). This is because F_m/K represents the static displacement of the mass if a static force of value F_m is applied, while A represents the dynamic displacement amplitude that occurs due to the periodic force of amplitude F_m. Note that the ratio R/R_c is known as the *damping ratio* δ, which can vary from 0 to ∞. If δ = 1.0, the damping is called *critical damping*. With vibration isolators, δ may be in the range from about 0.01 to 0.2.

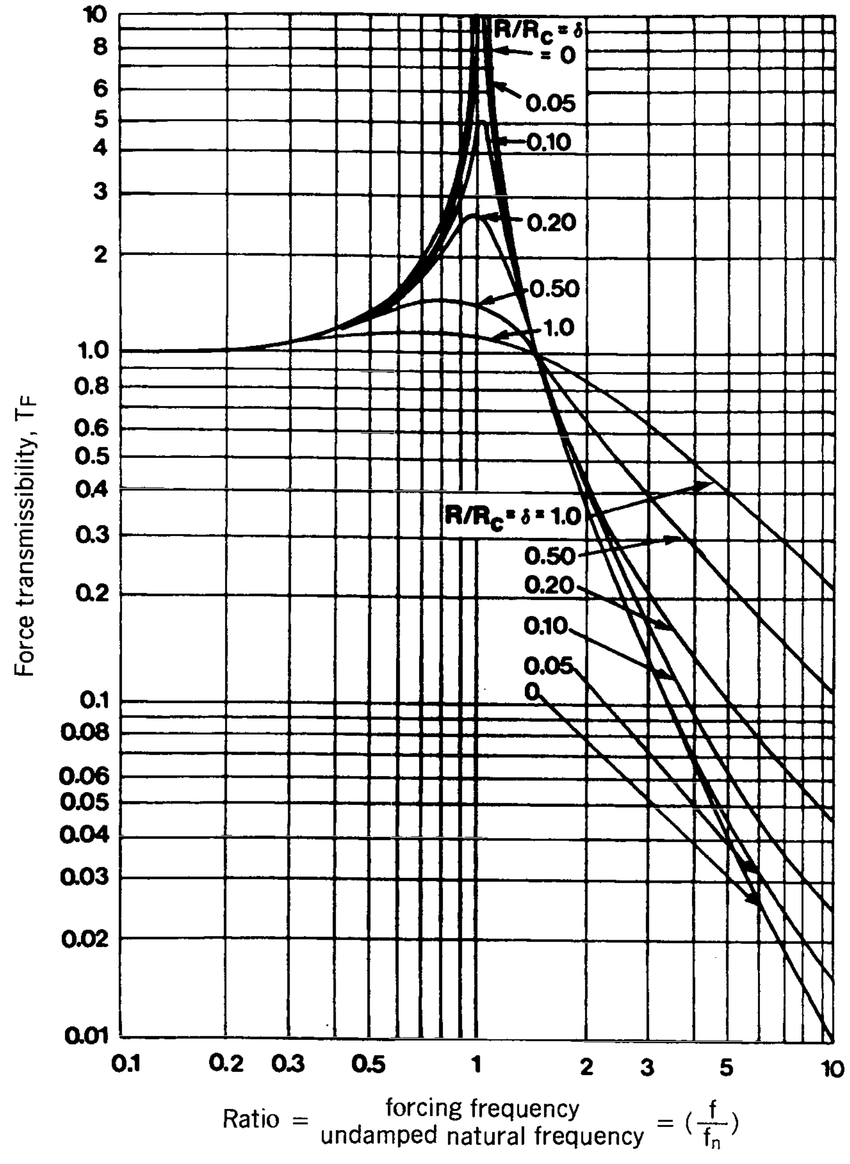


Fig. 4 Force transmissibility T_F for the rigid machine–isolator–rigid floor system.

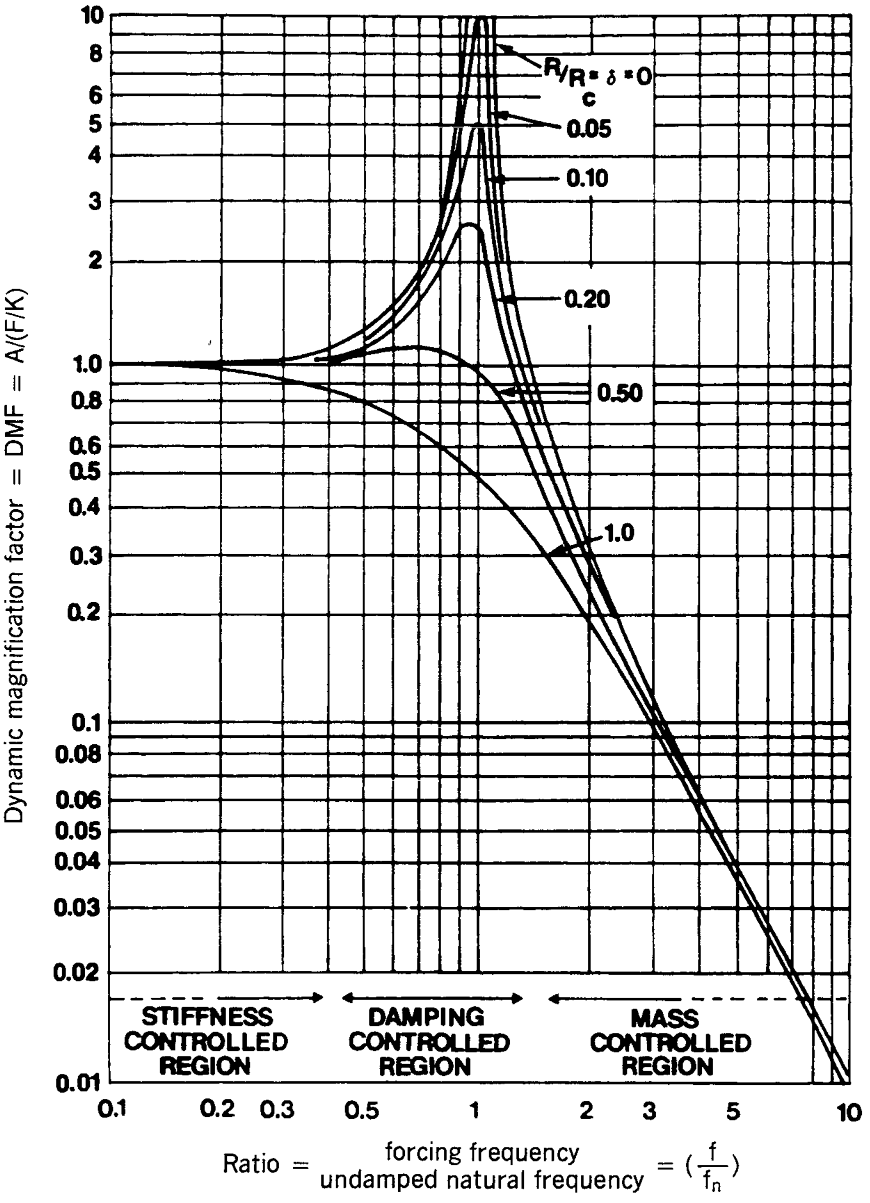


Fig. 5 Dynamic magnification factor DMF for the rigid machine–isolator–floor system.

Equations (2) and (3) are plotted in Figs. 4 and 5, respectively.

If the machine is run at the natural frequency f_n, we see from Fig. 4 that f/f_n is 1.0 and the force amplitude transmitted to the floor is very large, particularly if the damping in the isolator support is small. If the machine is operated much above the natural frequency, however, then the force amplitude transmitted to the floor will be very small.

Example 1. Suppose we wish to isolate the 120-Hz vibration of an electric motor. If we choose isolators so that the system has a natural frequency of 12 Hz, then the ratio $f/f_n = 10$. If the damping in the isolator system is $R/R_c = 0.1$, the force transmissibility will be only about 0.025, or 2.5%.

We define the efficiency η of the isolator as

$$\eta = (1 - T_F) \times 100\%. \qquad (4)$$

Thus in Example 1 the isolator efficiency is 97.5%.

In order to reduce the force transmissibility still further, we could use softer isolators and choose a still lower resonance frequency. There is some danger in doing this, however, because the static deflection of the machine will naturally increase if we use softer isolators. Since a large static deflection may be undesirable (it may interfere with the operation of the machine), this restricts the softness of the isolator and thus how low we can make the natural frequency f_n. The static deflection d produced in the isolator by the gravity force on the mass m is given by $d = mg/K$. We have already seen that the natural frequency is related to K and m by $f_n = (1/2\pi)\sqrt{K/m}$. Hence we can relate d and f_n:

$$d = \frac{g}{4\pi^2 f_n^2}, \tag{5}$$

where the static deflection d is given in centimetres (or inches) depending whether the acceleration of gravity g is 981 cm/s^2 or 386.4 in./s^2.

The relationship given in Eq. (5) has been plotted in Fig. 6. The greatest static deflection d that can be allowed from operational considerations should be chosen. This will then allow a determination of the lowest allowable f_n. It should be noted that excessive static deflection may interfere with the operation of the machine by causing alignment problems. Also, using isolators with a small vertical stiffness usually means that they will have a small horizontal stiffness, and there will be stability problems. Both these considerations limit the lowest allowable f_n.

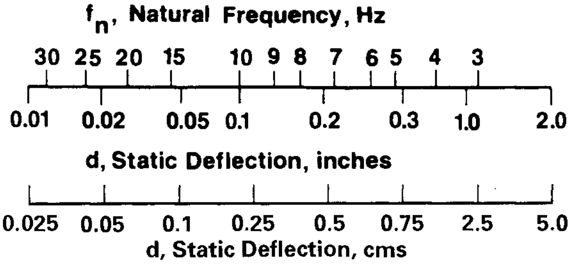


Fig. 6 Relationship between natural frequency f_n of machine–isolator–floor system and static deflection d of machine.

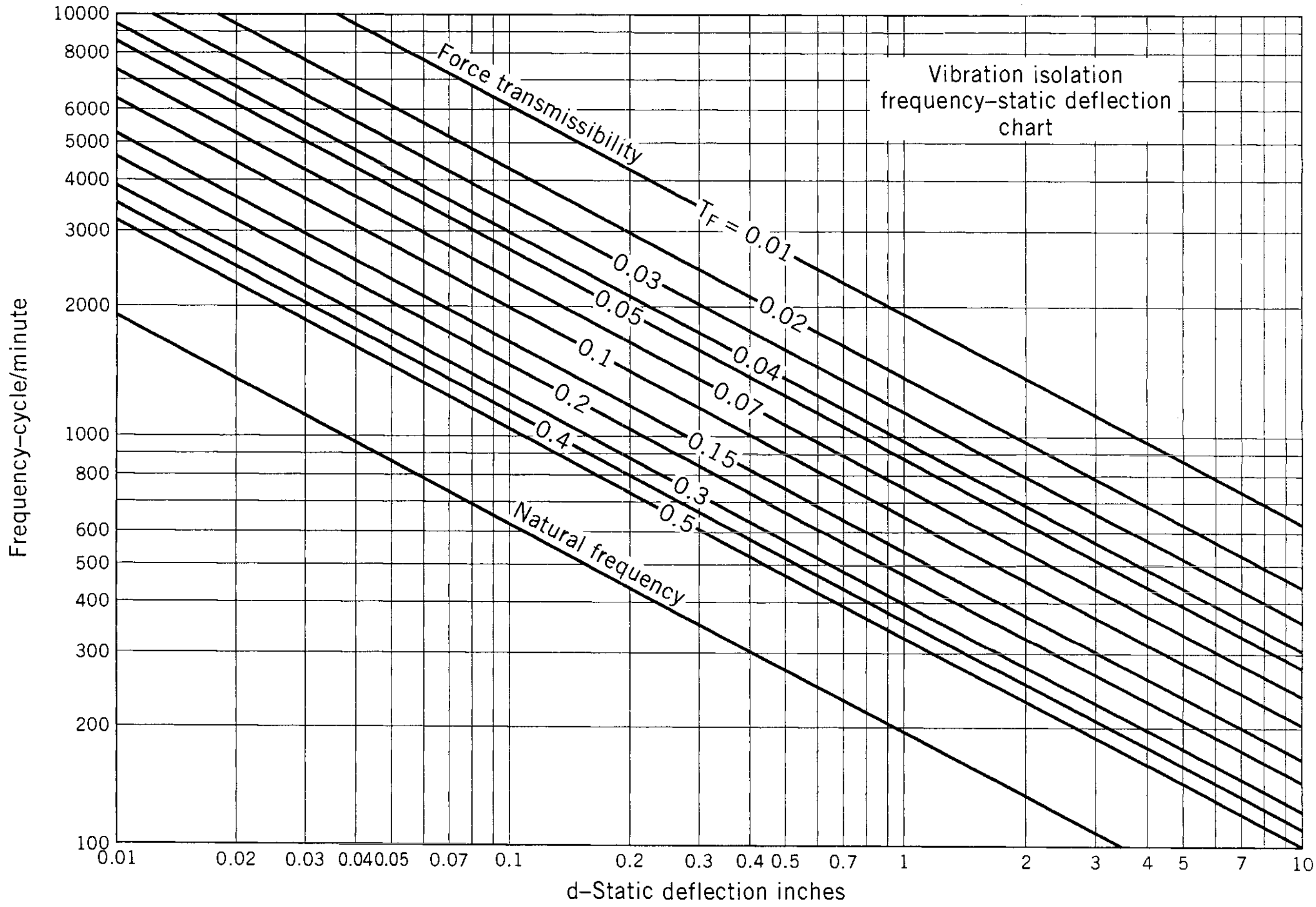


Fig. 7 Relationship between forcing frequency f_n, force transmissibility T_F and static deflection d of machine–isolator–floor system, without damping ($R = 0$).

Simple Design Procedure and Worked Example With springs, the damping is small (usually $\delta < 0.1$), and we may use Eqs. (2) and (3) and Figs. 4 and 5, assuming that $\delta = 0$. With $\delta = 0$ (or, equivalently, $R = 0$), Eqs. (2) and (5) give

$$T_F = \frac{F_t}{F_m} = \frac{1}{1 - 4\pi^2 f^2 d/g}. \tag{6}$$

This result is plotted in Fig. 7. The suggested procedure is as follows.

1. Establish the total weight of the machine and the lowest forcing frequency experienced.
2. From Fig. 7 select the force transmissibility allowable (this determines the static deflection, given the lowest forcing frequency).
3. From the spring constants given by the manufacturer, the machine weight, and the static deflection, choose the appropriate vibration isolator.

Example 2. An electric motor weighing 200 lb and a reciprocating compressor weighing 1000 lb are mounted on a common support. The motor runs at 2400 r/min and by a belt drives the compressor at 3000 r/min. The vertical force fed to the support is thought to be excessive. Choose six spring mounts to provide a force transmissibility not exceeding 5%.

1. The total machine weight is 1200 lb, and the lowest forcing frequency is 2400 r/min (or 40 Hz).
2. From Fig. 7, the static deflection required is 0.13 in.
3. Thus, since there are six spring isolators, each must support a weight of 200 lb. This requires a spring constant for each isolator of 200/0.13 = 1550 lb/in.

As a check on the calculation, we see from Fig. 6 that a static deflection of 0.13 in. requires a natural frequency of about 9 Hz. Thus, the ratio of forcing frequency to natural frequency $f/f_n = 40/9 \cong 4.5$. From Fig. 4, with zero damping, $T_F = 0.05$, which agrees with the design requirement.

Machine Vibration There are several other factors that should be considered in isolator design. First, we notice from Figs. 4 and 5 that when a machine is started or stopped, it will run through the resonance condition and then, momentarily, $f/f_n = 1$. When passing through the resonance condition, large vibration amplitudes can exist on the machine, and the force transmitted will be large, particularly if the damping is small. In order to provide small force transmissibility at the operating speed, small damping is required; however, to prevent excessive machine vibration and force transmission during stopping and starting, large damping is needed. These two requirements are conflicting. Fortunately, some forces, such as out-of-balance forces, are much reduced during starting and stopping. However, it is normal to provide a reasonable amount of damping ($\delta = 0.1$ to 0.2) in spring systems to reduce these starting and stopping problems. The severity of the machine vibration problem can be gauged from Fig. 8.

Use of Inertia Blocks Inertia blocks are normally made from concrete poured onto a steel frame. If the mass supported by the vibration isolators is increased by mounting a machine on an inertia block, the static deflection will be increased. If the isolator stiffness is correspondedly increased to keep the static deflection the same, then there is no change in the resonance frequency or the force transmissibility T_F. The use of an inertia block does, however, result in a reduced vibration amplitude of the machine mass. It also has additional advantages including (1) *improving stability* by providing vibration isolator support points that are farther apart, (2) *lowering the center of gravity* of the system, thus improving stability and reducing the effect of coupled modes and rocking natural frequencies, (3)

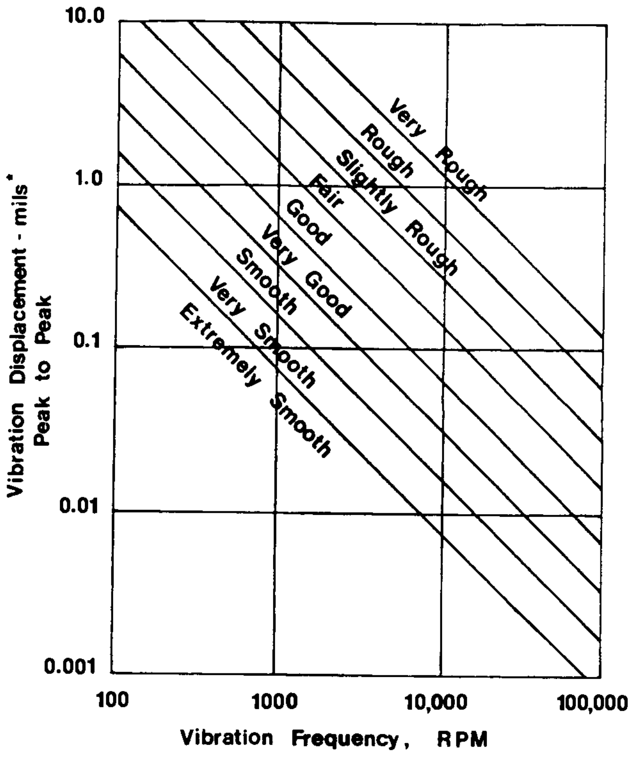


Fig. 8 Machine vibration severity chart showing peak to peak vibration (2A) mils (1 mil = 0.001 in. = 0.254 mm).

producing more even weight distribution for machines often enabling the use of symmetrical vibration isolation mounts, (4) *functioning as a local acoustic barrier* to shield the floor of an equipment room from the noise radiated from the bottom of the machine and reducing its transmission to rooms below, and (5) *reducing the effect on the machine of external forces* such as transient loads or torques caused by operation of motors or fans or rapid changes in machine load or speed.

Other Considerations The performance of vibration isolators in the high-frequency range ($f/f_n \gg 1$) is often disappointing. There are several possible reasons, but often they all result in an increase in force transmissibility T_F for $f/f_n \gg 1$. Usually, the reasons are because the simple single-degree-of-freedom model neglects various other effects mentioned now.[4]

(a) **Support Flexibility.** If the assumption that the support is rigid is poor, then the support will also deflect (e.g., a flexible floor). If the isolator stiffness is similar in magnitude to that of the supporting floor, then additional resonances in the floor–machine system will occur for $f/f_n > 1$. It is normal practice to choose isolators (assuming a rigid support or floor) to have a natural frequency well below the fundamental natural frequency of the floor itself. (If possible there should not be any machine-exciting frequencies in the frequency range 0.8 to 1.3 times the floor fundamental natural frequency.) The fundamental natural frequency of a wood floor is usually in the 20–30-Hz frequency range, while that of a concrete floor is in the 30–100-Hz range.

(b) **Machine Resonances.** Internal resonances in the machine structure will also increase the force transmissibility T_F in a similar manner to support flexibility. Increasing the stiffness and/or the damping of machine members can help to reduce this effect.

(c) **Standing Wave Effects.** Standing wave effects in the vibration isolators can significantly decrease their performance and increase force transmissibility for $f/f_n \gg 1$, particularly when the mass of the isolators becomes appreciable compared to the machine mass. This effect can be reduced by increasing the damping in the isolators. Springs particularly suffer from this problem because of their low inherent damping. Soft materials (such as felt or rubber) placed between the spring and the support can alleviate the problem. Elastometric isolators are less prone to this problem because of their higher damping.

(d) **Shock Isolation.** If the forces in the machine are impulsive in character (such as caused by repeated impacts), then the discussion for isolation of the single-degree-of-freedom model presented so far can still be used. It is normal practice to choose an isolator that provides a natural period T ($= 1/f_n$) much greater than the shock pulse duration but less than the period of repetition of the force.

Example 3. A nail-making machine cuts nails 5 times each second ($T = 0.2$ s). The shock pulse duration is approximately 0.015 s. What is a suitable choice of vibration isolators for the machine? Elastometric isolators that provide a natural period of vibration of 0.1 s (natural frequency of 10 Hz) would be a good choice. They have high damping.

4.2 Use of Sound Absorbing Materials

Sound absorbing materials have been found to be very useful in the control of noise. They are used in a variety of locations: close to sources of noise (e.g., close to sources in electric motors), in various paths, (e.g., above barriers) and sometimes close to a receiver (e.g. inside earmuffs). When a machine is operated inside a building, the machine operator usually receives most sound through a direct path, while people situated at greater distances receive sound mostly through reflections (see Fig. 9).

The relative contributions of the sound reaching people at a distance through direct and reflected paths are determined by how well the sound is reflected and absorbed by the walls in the building. The *absorption coefficient* of a material $\alpha(f)$, which is a function of frequency, has already been defined in Chapter 1 of Volume 1:

$$\alpha(f) = \frac{\text{sound intensity absorbed}}{\text{sound intensity incident}}. \tag{7}$$

Thus $\alpha(f)$ is the fraction of incident sound intensity that is absorbed, and it can vary between 0 and 1. Materials that have a high value of α are usually fibrous or porous. Fibrous materials include those made from natural or artificial fibers including glass fibers. Porous materials made from open-celled polyurethane foams are widely used.

It is believed that there are two main mechanisms by which the sound is absorbed in materials: (1) viscous dissipation of energy by the sound waves as they propagate in the narrow channels of the material and (2) energy losses caused by friction as the fibers of the material rub together under the influence of the sound waves. These mechanisms are illustrated in Fig. 10. In

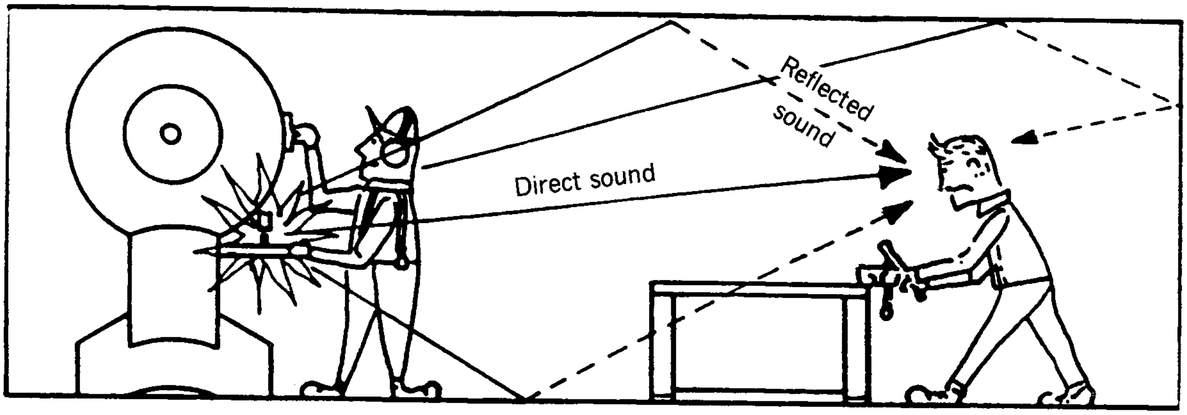


Fig. 9 Paths of direct and reflected sound emitted by a machine in a building.

both mechanisms sound energy is converted into heat. The sound absorption coefficient of most acoustic materials increases with frequency (see Fig. 11), and coefficients of some common sound absorbing materials and construction materials are shown in Tables 2 and 3. The noise reduction coefficient NRC of a sound absorbing material is defined as the average of the absorption coefficients at 250, 500, 1000 and 2000 Hz (rounded off to the nearest multiple of 0.05).

In the case of machinery used in reverberant spaces such as factory buildings, the reduction in sound pressure level L_p in the reverberant field caused by the addition of sound absorbing material placed on the walls or under the roof (see Fig. 12) can be estimated for a source of sound power level L_w from the so-called room equation:

$$L_p = L_w + 10\log\left(\frac{D}{4\pi r^2} + \frac{4}{R}\right), \qquad (8)$$

where the *room constant* $R = S\overline{\alpha}/(1 - \overline{\alpha})$, $\overline{\alpha}$ is the surface average absorption coefficient of the walls, D is the source directivity (see Chapter 1), and r is the distance in metres from the source. The surface average absorption coefficient $\overline{\alpha}$ may be estimated from

$$\overline{\alpha} = \frac{S_1\alpha_1 + S_2\alpha_2 + S_2\alpha_2 + \cdots}{S_1 + S_2 + S_3 + \cdots}, \qquad (9)$$

where $S_1, S_2, S_3, \ldots,$ are the surface areas of material with absorption coefficients $\alpha_1, \alpha_2, \alpha_3, \ldots$. For the sus-

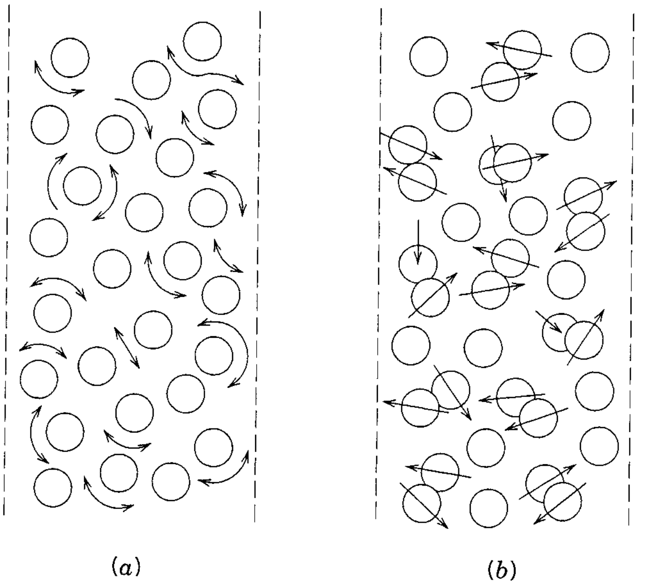

(*a*) (*b*)

Fig. 10 The two main mechanisms believed to exist in sound absorbing materials: (*a*) viscous losses in air channels and (*b*) mechanical friction caused by fibers rubbing together.

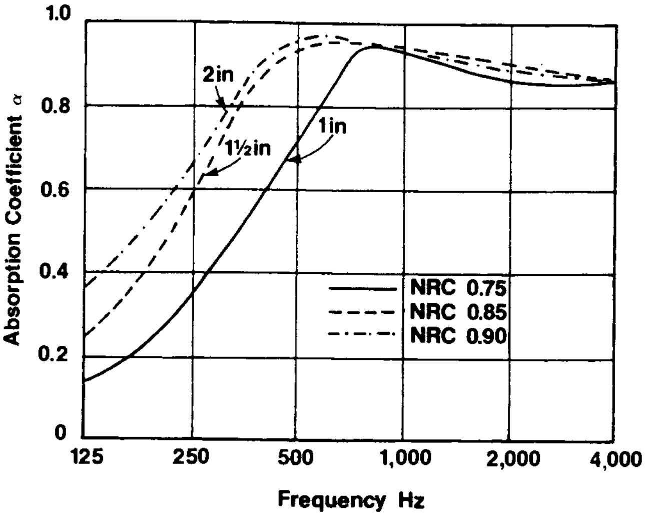


Fig. 11 Sound absorption coefficient α and noise reduction coefficient NRC for typical fiberglass formboard.

TABLE 2 Sound Absorption Coefficients of Common Acoustic Materials

Materials	Frequency (Hz)					
	125	250	500	1000	2000	4000
Fibrous glass (typically 4 lb/cu ft) hard backing						
1 in. thick	0.07	0.23	0.48	0.83	0.88	0.80
2 in. thick	0.20	0.55	0.89	0.97	0.83	0.79
4 in. thick	0.39	0.91	0.99	0.97	0.94	0.89
Polyurethane foam (open cell)						
$\frac{1}{4}$ in. thick	0.05	0.07	0.10	0.20	0.45	0.81
$\frac{1}{2}$ in. thick	0.05	0.12	0.25	0.57	0.89	0.98
1 in. thick	0.14	0.30	0.63	0.91	0.98	0.91
2 in. thick	0.35	0.51	0.82	0.98	0.97	0.95
Hairfelt						
$\frac{1}{2}$ in. thick	0.05	0.07	0.29	0.63	0.83	0.87
1 in. thick	0.06	0.31	0.80	0.88	0.87	0.87

TABLE 3 Sound Absorption Coefficient $\alpha(f)$ of Common Construction Materials

Material	Frequency f (Hz)					
	125	250	500	1000	2000	4000
Brick						
Unglazed	0.03	0.03	0.03	0.04	0.04	0.05
Painted	0.01	0.01	0.02	0.02	0.02	0.02
Concrete block, painted	0.10	0.05	0.06	0.07	0.09	0.03
Concrete	0.01	0.01	0.015	0.02	0.02	0.02
Wood	0.15	0.11	0.10	0.07	0.06	0.07
Glass	0.35	0.25	0.18	0.12	0.08	0.04
Gypsum board	0.29	0.10	0.05	0.04	0.07	0.09
Plywood	0.28	0.22	0.17	0.09	0.10	0.11
Soundblox concrete block						
Type A (slotted), 6 in.	0.62	0.84	0.36	0.43	0.27	0.50
Type B, 6 in.	0.31	0.97	0.56	0.47	0.51	0.53
Carpet	0.02	0.06	0.14	0.37	0.60	0.66

pended absorbing panels shown in Fig. 12, both sides of the panel are normally included in the surface area calculation.

If the sound absorption is increased, then from Eq. (8) the change in sound pressure level ΔL in the reverberant space (beyond the critical distance r_c) (see Chapter 1 Volume 1) is

$$\Delta L = L_{p1} - L_{p2} = 10 \log \frac{R_2}{R_1}. \qquad (10)$$

If $\overline{\alpha} \ll 1$, then the reduction in sound pressure level (sometimes called the *noise reduction*) is given by

$$\Delta L \approx 10 \log \frac{S_2 \overline{\alpha}_2}{S_1 \overline{\alpha}_1}, \qquad (11)$$

where S_2 is the total surface area of the room walls, floor, and ceiling and any suspended sound absorbing material, $\overline{\alpha}_2$ is the average sound absorption coefficient of these surfaces after the addition of sound absorbing material and S_1 and $\overline{\alpha}_1$ are the area and the average sound absorption coefficient before the addition of the material.

Example 4. A machine source operates in a building of dimensions 30 m × 30 m with a height of 10 m. Suppose the average absorption coefficient is $\overline{\alpha} = 0.02$ at 1000 Hz. What would be the noise reduction in the reverberant field if 100 sound absorbing panels with dimensions 1 m × 2 m each with an absorption coefficient of $\alpha = 0.8$ at 1000 Hz were suspended from the ceiling (assume both sides absorb sound)? The room surface area $= 2(900) + 4(300) = 3000\ \text{m}^2$, therefore $R_1 = (3000 \times 0.02)/0.98 = 60/0.98 = 61.2$ sabins (m^2). The new aver-

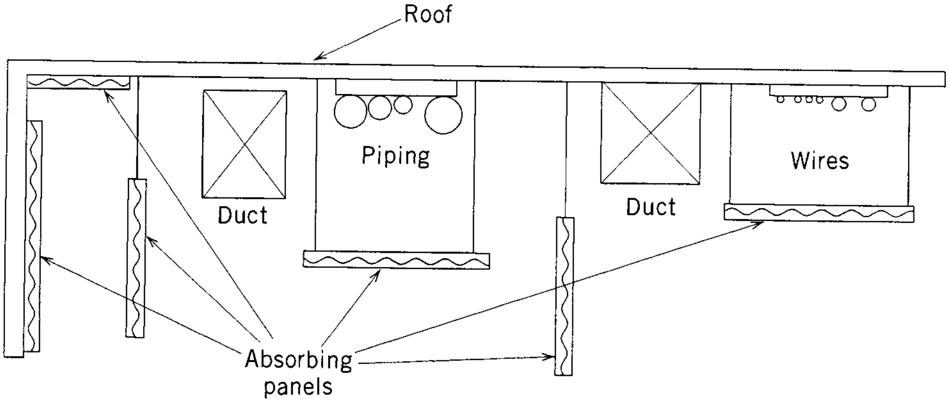


Fig. 12 Sound absorbing material placed on the walls and under the roof and suspended as panels in a factory building.

age absorption coefficient $\overline{\alpha}_2 = [3000 \times 0.02 + 200 \times 2 \times 0.8]/3400 = 60 + 320 = 380/3400 = 0.11$. The new room constant is $(3400 \times 0.11)/0.89 = 420$ sabins (m^2). Thus from Eq. (10) the predicted noise reduction $\Delta L = 10 \log (420/61.2) = 8.3$ dB. This calculation may be repeated at each frequency for which absorption coefficient data are available. It is normal to assume that about 10 dB is the practical limit for the noise reduction that can be achieved by adding sound absorbing material in industrial situations.

4.3 Acoustic Enclosures

Acoustic enclosures are used wherever containment or encapsulation of the source or receiver are a good, cost-effective, feasible solution. They can be classified in four main types: (1) large loose-fitting or room-size enclosures in which complete machines or personnel are contained, (2) small enclosures used to enclose small machines or parts of large machines, (3) close-fitting enclosures that follow the contours of a machine or a part, and (4) wrapping or lagging materials often used to wrap pipes, ducts, or other systems.

The performance of such enclosures can be defined in three main ways[4]: (1) *noise reduction* (NR), the difference in sound pressure levels between the inside and outside of the enclosure,* (2) *transmission loss* (TL, or equivalently the *sound reduction index*), the difference between the incident and transmitted sound intensity levels for the enclosure wall, and (3) *insertion loss* (IL), the difference in sound pressure levels at the receiver point *without* and *with* the enclosure wall in place. Enclosures can either be complete or partial (in which some walls are removed for convenience or accessibility). Penetrations are also often necessary to provide access or cooling.

The transmission coefficient τ of a wall may be defined as

$$\tau = \frac{\text{sound intensity transmitted by wall}}{\text{sound intensity incident on wall}}, \qquad (12)$$

and this is related to the transmission loss TL (or sound reduction index) by

$$\text{TL} = 10 \log \frac{1}{\tau}. \qquad (13)$$

If the sound fields can be considered to be reverberant on the two sides of a complete enclosure, then

$$\text{NR} = L_{p1} - L_{p2} = TL + 10 \log \frac{A_2}{S_e}, \qquad (14)$$

where L_{p1} and L_{p2} are the sound pressure levels on the transmission and receiving sides of the enclosure, $A_2 = S_2\overline{\alpha}_2$ is the absorption area in square metres or feet in the receiving space material where $\overline{\alpha}_2$ is the average absorption coefficient of the absorption material in the receiving space averaged over the area S_2, and S_e is the enclosure surface area in square metres or feet.

Equation (14) can be used to determine the transmission loss. TL of a partition of surface area S_e placed between two isolated reverberation rooms in which a noise source creates L_{p1} in the source room and this results in L_{p2} in the receiving room. Also Eq. (14) can be used to design a personnel enclosure. If the enclosure is located in a factory building in which the reverberant level is L_{p1}, then the enclosure wall TL and interior absorption area A_2 can be chosen to achieve an acceptable value for the interior sound pressure level L_{p2}. In the case that the surface area S_2 of the interior absorbing material $S_2 = S_e$, the enclosure surface area, then Eq.

---

*Note: The *noise reduction* (NR) is not the same as the *noise reduction* discussed in Section 4.2.

(14) simplifies to:

$$\mathrm{NR} = \mathrm{TL} + 10 \log \bar{\alpha}_2 \qquad (15)$$

We see that generally the noise reduction NR achieved is less than the transmission loss TL. We note that when $\bar{\alpha}_2$, the average absorption coefficient of the absorbing material in the receiving space approaches 1, then NR → TL (the expected result), although when $\bar{\alpha}_2 \rightarrow 0$ then the theory fails.

Example 5. If the reverberant level in a factory space is 90 dB in the 1000 Hz one-third octave band, what values of TL and $\bar{\alpha}$ should be chosen to insure that the interior level inside a personnel enclosure is below 60 dB? Assuming that $S_2 = S_e$, then if TL is chosen as 40 dB and $\bar{\alpha} = 0.1$, NR = 40 − 10 log 0.1 = 40 − 10 = 30 dB, and $L_{p2} = 60$ dB. If $\bar{\alpha}$ is increased to 0.2, then NR = 40 − 10 log 0.2 = 40 − 7 = 33 and $L_{p2} = 57$ dB meeting the requirement. Note that in general, since TL varies with frequency (see Eq. 18) then this calculation would have to be repeated for each one-third octave band center frequency of interest. At low frequency, since large values of TL and $\bar{\alpha}$ are difficult to achieve, it may not be easy to obtain large values of NR.

When an enclosure is used to contain a source, it works by reflecting the sound field back towards the source causing an increase in the sound pressure level inside the enclosure. From energy considerations the insertion loss would be zero if there were no acoustical absorption inside. The buildup of sound energy inside the enclosure, however, can be reduced significantly by the placement of sound absorbing material inside the enclosure. It is also useful to place sound absorbing materials inside personnel protective booths for similar reasons. In general, it is difficult to predict the insertion loss of an enclosure. For one installed around a source with a considerable amount of absorbing material used inside to prevent *any* interior reverberant sound energy buildup, then from energy considerations:

$$\mathrm{IL} \approx \mathrm{TL}$$

if the receiving space is quite absorbent and if IL and TL are wide-frequency-band averages (e.g., at least one octave). If insufficient absorbing material is placed inside the enclosure, the sound pressure level will build up inside the enclosure by multiple reflections, and the enclosure effectiveness will be degraded. From energy considerations it is obvious that if there is no sound absorption inside ($\bar{\alpha} = 0$), the enclosure will be useless, and its insertion loss IL will be zero.

An estimate of the insertion loss of intermediate cases $0 < \bar{\alpha} < 1$ can be obtained by assuming that the sound field inside the enclosure is reverberant and that the interior surface of the enclosure is lined with absorbing material of average absorption coefficient $\bar{\alpha}$. With the assumptions that (1) the average absorption coefficient in the room containing the noise source is not greater than 0.3 (which is true for most reverberant factory or office spaces), (2) the noise source does not provide direct mechanical excitation to the enclosure, and (3) the noise source occupies less than about 0.3 to 0.4 of the enclosure volume, then it may be shown that the insertion loss of a loose-fitting enclosure made to contain a noise source situated in a reverberant room is[5]

$$\mathrm{IL} = L_{p1} - L_{p2} = \mathrm{TL} + 10 \log \frac{A_e}{S_e}$$

where L_{p1} is the reverberant level in the room containing the noise source (without the enclosure), L_{p2} is the reverberant level at the same point (with the enclosure), A_e is the absorption area inside the enclosure $S_i\bar{\alpha}_i$, and $\bar{\alpha}_i$ = surface average absorption coefficient of this material. In the case that the surface area of the enclosure $S_e = S_i$, the area of the interior absorbing material, then Eq. (16) simplifies to

$$\mathrm{IL} = \mathrm{TL} + 10 \log \bar{\alpha}_i. \qquad (17)$$

We see that in general the insertion loss of an enclosure containing a source is less than the transmission loss TL. We note that when $\bar{\alpha}_i$ the average absorption coefficient of the internal absorbing material approaches 1, the IL → TL (the expected result), although when $\bar{\alpha} \rightarrow 0$ then this theory fails.

Example 6. If the sound pressure level L_{p1} in a reverberant factory space caused by a machine source in the one-third octave band is 85 dB, what values of TL and $\bar{\alpha}$ should be chosen to ensure that the reverberant level is reduced to be below 60 dB? Assuming that $S_i = S_e$, then if TL is chosen to be 30 dB and $\bar{\alpha} = 0.1$, IL = 30 −10 log 0.1 = 30 − 10 = 20 dB and $L_{p2} = 65$ dB. If $\bar{\alpha}$ is increased to 0.2, then IL = 30 −10 log 0.2 = 30 − 7 = 23 dB and $L_{p2} = 62$ dB. If $\bar{\alpha}$ is increased to 0.4, then IL = 30 −10 log 0.4 = 30 − 4 = 26 dB and $L_{p2} = 59$ dB, thus meeting the requirement. As in Example 5, we note that since TL varies with frequency f (see Eq. 18) this calculation should be repeated at each frequency of interest. At low frequency, since large values of TL and $\bar{\alpha}$ are hard to obtain, it may be difficult to obtain large values of IL.

Thus the transmission loss TL can be used as a rough guide to the IL of a sealed enclosure only if allowance is made for the sound absorption inside the enclosure. The transmission loss of an enclosure is normally dominated by the mass/unit area m of the enclosure walls (except in the coincidence-frequency region). This is because

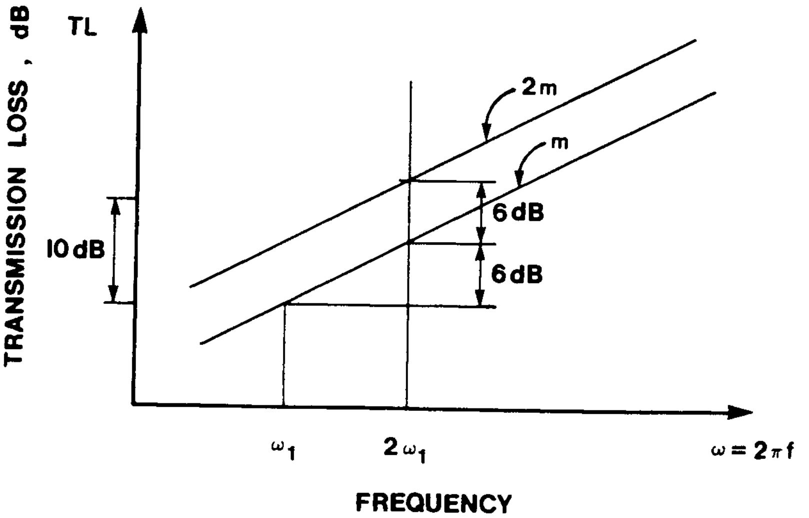


Fig. 13 Variation of mass law transmission loss TL of a single panel (see Eq. 18).

Enclosures

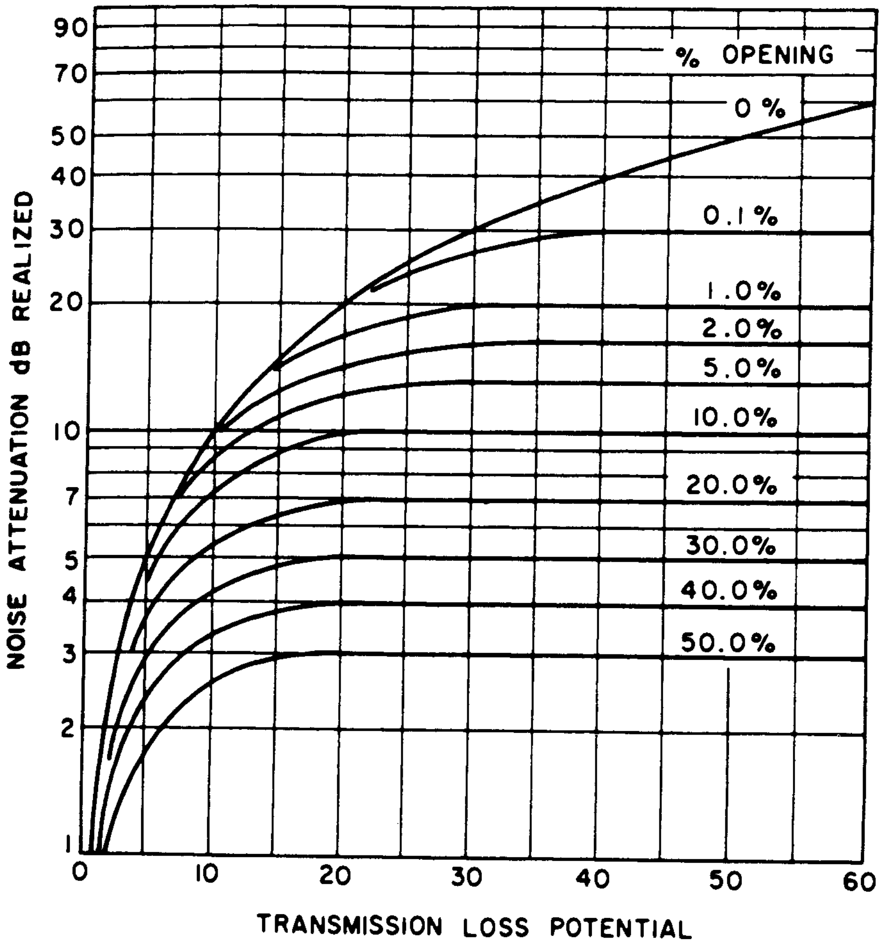


Fig. 14 Transmission loss of enclosure walls with holes as a function of transmission loss TL of enclosure wall without holes and percentage open area of holes.

when the stiffness and damping of the enclosure walls are unimportant, the response is dominated by inertia $m(2\pi f)$ where f is the frequency in hertz. The transmission loss of an enclosure wall for sound arriving from all angles is approximately

$$\mathrm{TL} = 20\log(mf) - C \quad \mathrm{dB} \tag{18}$$

where m is the surface density (mass/unit area) of the enclosure walls and $C = 47$ if the units of m are kg/m^2 and $C = 34$ if the units are lb/ft^2. Figure 13 shows that the transmission loss of a wall (given by Eq. 18) theoretically increases by 6 dB for doubling of frequency or for doubling of the mass/unit area of the wall. If some areas of the enclosure walls have a significantly poorer TL than the rest, then this results in a degradation in the overall transmission loss of the enclosure. Where the enclosure surface is made of different materials (e.g., wood walls and window glass), then the average transmission loss $\mathrm{TL}_{\mathrm{ave}}$ of the composite wall is given by

$$TL_{\mathrm{ave}} = 10\log\frac{1}{\bar{\tau}}, \tag{19}$$

where the average transmission coefficient $\bar{\tau}$ is

$$\bar{\tau} = \frac{S_1\tau_1 + S_2\tau_2 + \cdots + S_n\tau_n}{S_1 + S_2 + S_3 + \cdots + S_n}, \tag{20}$$

where τ_i = transmission coefficient of the ith wall element and S_i = surface area of ith wall element (m^2). If holes or leaks occur in the enclosure walls and if the TL of the holes is assumed to be 0 dB, then Fig. 14 shows the degradation in the average TL of the enclosure walls. If the penetrations in the enclosure walls are lined with absorbing materials as shown in Fig. 15, then the degradation in the enclosure TL is much less significant.

Enclosures are available from manufacturers in a wide variety of ready-made modular panels (see Fig. 16).

Close-fitting Enclosures If the noise source occupies no more than about one third of the volume of a sealed enclosure, then the simple theory described by Eqs. (16) and (17) can be used. However, in many cases when machines are enclosed it is necessary to locate the enclosure walls close to the machine surfaces, so that the

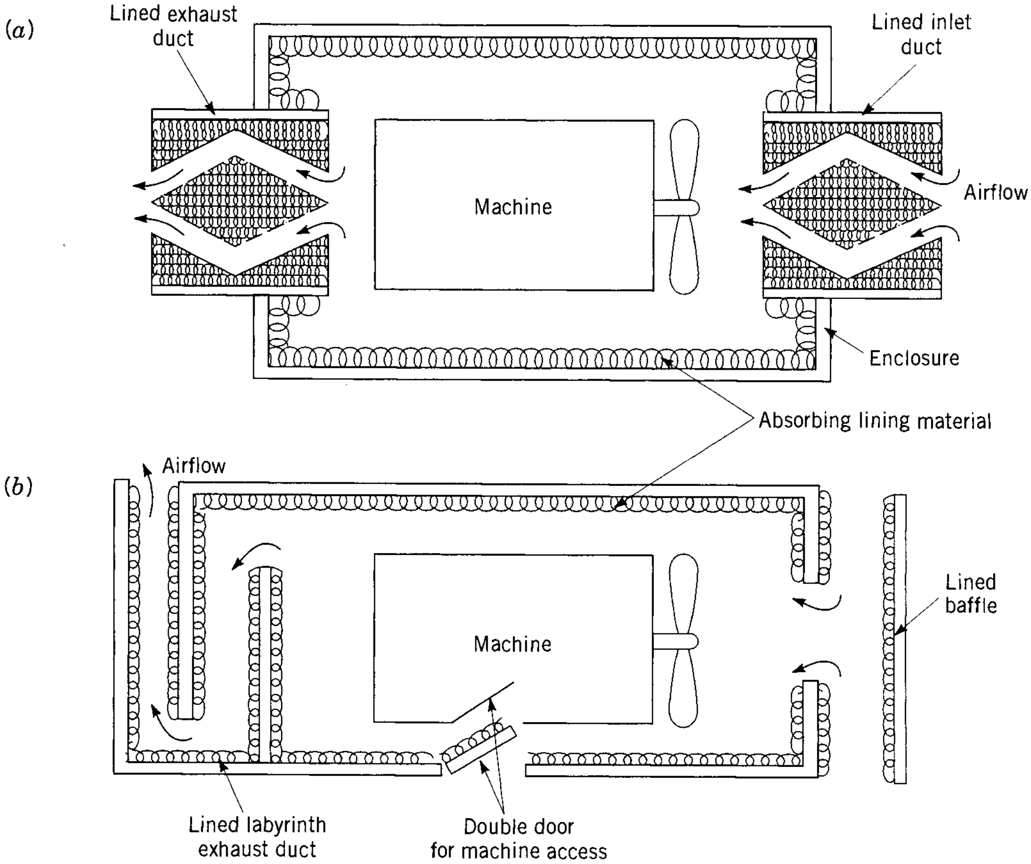


Fig. 15 Enclosures with penetrations (for cooling) lined with absorbing materials: (*a*) lined ducts, (*b*) lined baffles with double door access provided to interior of machine.

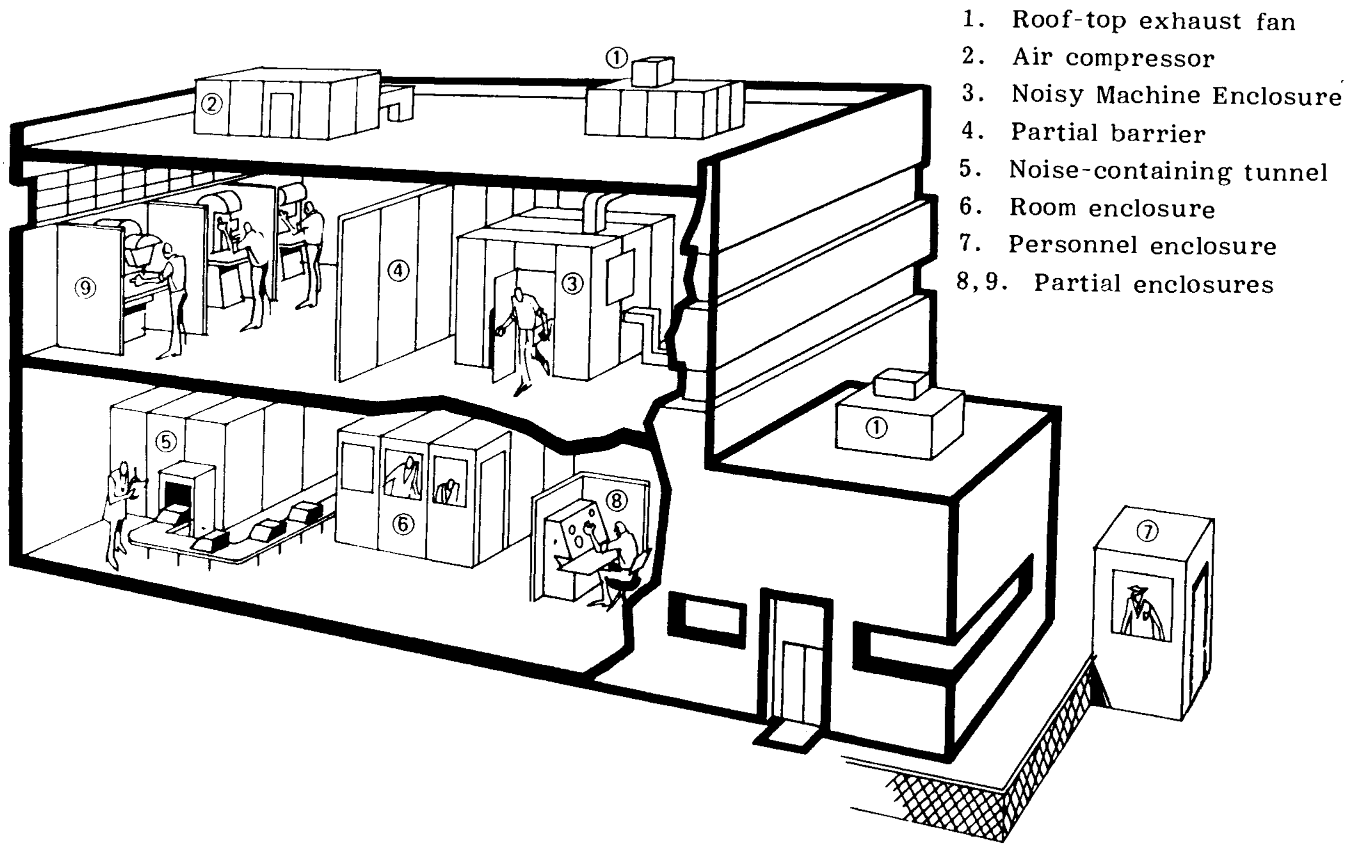


Fig. 16 Ready made modular materials used to make enclosures and barriers in a factory building. (Courtesy of Lord Corporation, Erie, PA.)

resulting air gap is small. Such enclosures are termed *close-fitting enclosures*. In such cases the sound field inside the enclosure is not reverberant or diffuse and the theory discussed so far can be used for a first approximation of the insertion loss of an enclosure. There are several effects that occur with close-fitting enclosures.

First, if the source has a low internal impedance, then in principle the close-fitting enclosure can "load" it so that it produces less sound power. However, in most machinery noise problems, the internal impedance of the source is high enough to make this effect negligible. Second, and more importantly, are the reductions in the insertion loss IL that occur at certain frequencies (when the enclosure becomes "transparent"). See the frequencies f_0 and f_{sw} in Fig. 17. When an enclosure is *close-fitting*, then to a first approximation the sound waves approach the enclosure walls at normal incidence instead of random incidence. When the air gap is small, then a resonant condition at frequency f_0 occurs where the enclosure wall mass is opposed by the wall and air gap stiffness. In addition standing wave resonances can occur in the air gap at frequencies f_{sw}. These resonances can be reduced by the placement of sound absorbing material in the air gap.[6,7] Jackson has produced simple theoretical models for close-fitting enclosures which assume a uniform air gap.[6,7] However, in practice the air gap varies with real enclosures and these simple theoretical mod-

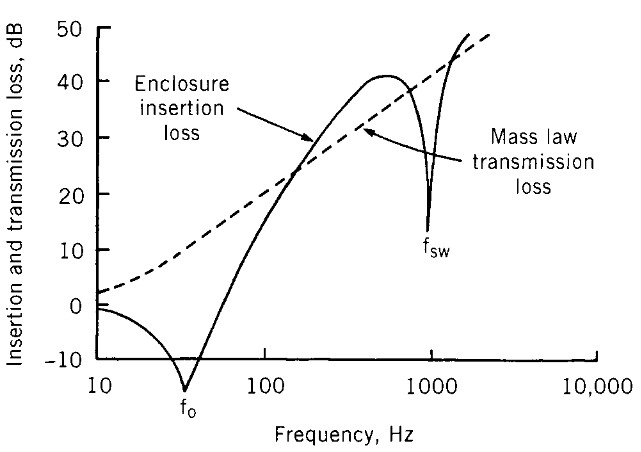


Fig. 17 Insertion loss and transmission loss of an enclosure, (mass/unit area = 16 kg/m^2, air gap = 16 cm) (from Ref. 6).

els and some later ones can only be used to give some guidance of the insertion loss to be expected in practice. Finite element and boundary element approaches can be used to make insertion loss predictions for close-fitting enclosures with complicated geometries.

4.4 Barriers

An obstacle placed between a noise source and a receiver is termed a *barrier* or *screen*. As explained in Chapter 1 when a sound wave approaches the barrier, some of the sound wave is reflected and some is transmitted past (see Fig. 18). At high frequency barriers are quite effective, and a strong acoustic "shadow" is cast. At low frequency (when the wavelength, can equal or exceed the barrier height), the barrier is less effective, and some sound is diffracted into the shadow zone. Indoors, barriers are usually partial walls. Outdoors, the use of walls, earth berms, and even buildings can protect residential areas from traffic and industrial noise sources. Empirical charts are available to predict the attenuation of a barrier.[4] Figure 19 shows the *insertion loss* or reduction in sound pressure level expected after installation of a semi-infinite barrier in free space between a source and receiver. If barriers are used inside buildings, their performance is often disappointing because sound can propagate into the shadow zone by multiple reflections. To produce acceptable attenuation it is important to suppress these reflections by the use of sound absorbing material, particularly on ceilings just above the barrier. In Fig. 19, N is the dimensionless Fresnel number related to the shortest distance over the barrier $d_1 + d_2$ and the straight line distance d between the source S and receiver R:

$$N = 2(d_1 + d_2 - d)/\lambda, \qquad (21)$$

where λ is the wavelength and d_1 and d_2 are shown in Fig. 19.

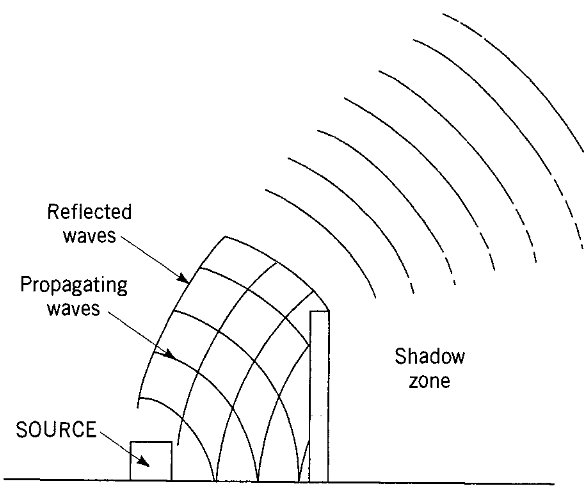


Fig. 18 Sound waves reflected and diffracted by barrier and acoustic shadow zone.

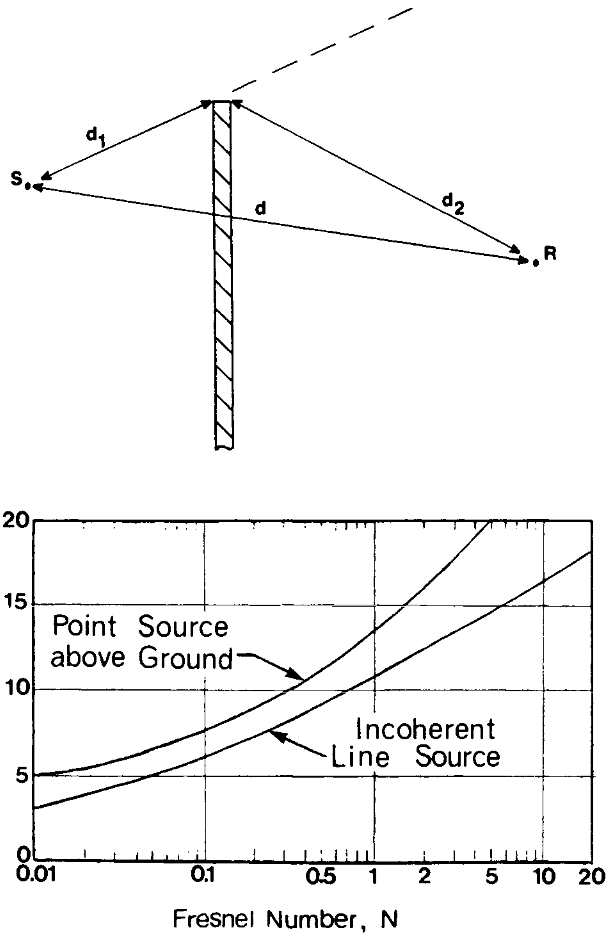


Fig. 19 Attenuation of a barrier as a function of Fresnel number N for point and incoherent line source.

4.5 Damping Materials

Load-bearing and non–load-bearing structures of a machine (panels) are excited into motion by mechanical machine forces resulting in radiated noise. Also, the sound field inside an enclosure excites its walls into vibration. When resonant motion dominates the vibration, the use of damping materials can result in significant noise reduction. In the case of machinery enclosures, the motion of the enclosure walls is normally mass controlled (except in the coincidence-frequency region), and the use of damping materials is often disappointing. Damping materials are often effectively employed in machinery in which there are impacting parts since these impacts excite resonant motion.

The mechanisms involved in damping all involve the conversion of mechanical energy into heat and include friction (rubbing) of parts, air pumping at joints, sound

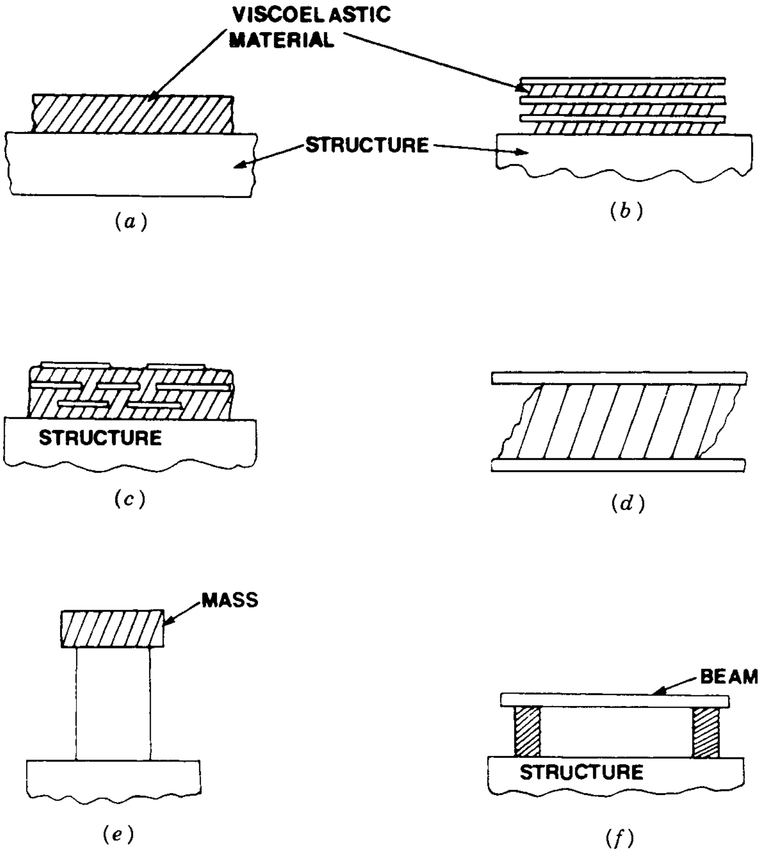


Fig. 20 Different ways of using vibration damping materials: (*a*) Free layer, (*b*) multiple constrained layer, (*c*) multi-layer spaced treatment, (*d*) sandwich panel, (*e*) tuned damper, (*f*) resonant beam damper.

radiation, viscous effects, eddy currents, and magnetic and mechanical hysteresis. Rubbery, plastic, and tarry materials usually possess high damping. During compression, expansion or shear, these materials store elastic energy. When the motion is reversed, some viscous flowing occurs, and the energy stored is converted into heat. The damping properties of such materials are temperature dependent. Damping materials can be applied to structures in a variety of ways. Figure 20 shows some common ways of applying damping materials and systems to structures.

5 MACHINERY NOISE SOURCES

Many different machines are used in appliances, in transportation vehicles, and in industry. But some components such as bearings, gears, and fans are used in many more complex machines. It is impossible to discuss every type of machine here. The discussion following will of necessity be brief and concern some of the most common machine components. Several references discuss machinery noise mechanisms and sources in more detail than is possible here.[8–14]

5.1 Bearings

There are two main types of bearings: (1) *rolling contact* and (2) *sliding contact*.[15,16] Rolling contact bearings are more commonly used, but sliding contact bearings are usually quieter than rolling contact bearings, if properly manufactured and maintained. Proper lubrication is essential for both rolling and sliding contact bearings.

Rolling contact bearings consist of the rolling elements contained between the *inner* and *outer raceways*. The rolling elements are normally kept from touching each other by a *cage*. The rolling elements may be spherical, cylindrical, tapered, or barrel-shaped.[15] Figure 21 shows a bearing with spherical rolling elements. The noise made by a rolling contact bearing is normally caused by vibration from two main sources: (1) rotation of bearing elements and (2) resonances in the elements, raceways, or cage. Shahan and Kamperman have identified seven discrete frequencies (and their harmonics) that

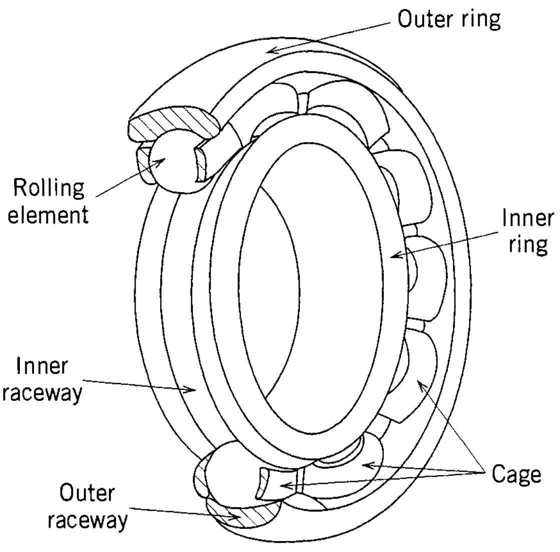


Fig. 21 Bearing with spherical rolling elements.

are related to bearing geometry and rotational speed.[13] The first is the shaft rotational frequency f_s, where

$$f_s = N/60, \qquad (22)$$

where N is shaft rotational speed rpm.

The other six frequencies are related to the shaft frequency f_s by factors that depend on the roller diameter, the pitch diameter of the bearing, the contact angle between the roller element and the raceway, and the number of rolling elements. Manufacturing imperfections cause bearing noise, which can be increased further by wear. Even a perfect bearing will make a noise when loaded.[9] If a rolling bearing is manufactured to a higher grade of precision (smaller tolerance), then it normally becomes quieter (and more expensive). (Classes of tolerance are specified in ISO Standard 492-1986. Methods to test bearings for noise and vibration are given in ANSI/AFBMA Standard 13-1970. The (American) Military Specification MIL-B-17913D defines permissible vibration limits for bearings.)

Sliding contact bearings are of three main types: (1) journal, (2) thrust, and (3) guide. Journal bearings are cylindrical in shape and allow rotation (see Fig. 22). Thrust bearings are used to prevent motion along a shaft axis, while guide bearings are normally used for motion of a part in one direction without rotation (for example, a piston sliding in a cylinder of an internal combustion engine). When shaft rotation occurs with a journal bearing, the shaft rides on a film of lubricant. Under some conditions, however, this film can break down causing metal-to-metal contact and noise and vibration. A well-known instability condition known as *oil whirl* can occur causing noise at a frequency of about half of the shaft rotational speed. This is because the average speed of the lubricant film is about half of the shaft speed. Oil-whirl noise can be magnified if a shaft resonance frequency occurs near to half the shaft rotational speed. To minimize sliding-bearing noise, proper attention should be paid to lubricant viscosity, pressure, alignment, and structural stiffness. References 9, 13, 14, and 16 contain more detailed discussion on sliding-bearing noise.

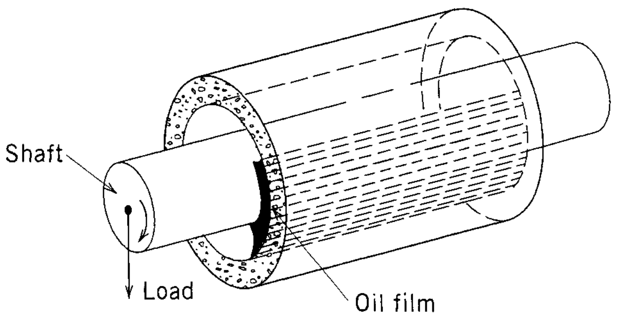


Fig. 22 Sliding contact bearing.

5.2 Gears

Most modern gear teeth have an involute profile, although some have circular-arc profiles.[8,9] Figure 23*a* shows some of the terms used with involute gears and Fig. 23*b* shows two parallel-axis spur-type gears meshing together.

There are several different types of gears in common use, as shown in Figure 24. They are of two main classes, either having: (1) parallel axes (spur, helical) or (2) nonparallel axes (straight bevel, spiral bevel, hypoid). These spur gears and straight-bevel gears are usually the noisiest, while the helical spiral-bevel gears are usually the quietest of their respective classes. This is because the load between gear wheel teeth is transferred more gradually with helical and spiral-bevel gears rather than more abruptly with spur and straight-bevel gears. Gear noise can arise from a variety of causes. As the gear teeth mesh, a pulsating force occurs at the gear-tooth meshing frequency f_m and its harmonics. Harmonics are present because the pulsating force on the gear teeth is not purely sinusoidal. The strength of the harmonics depends on the force pulse shape and also on other impulsive forces caused by tooth deformation, machinery errors, bearing misalignment, pinion wheel deformation, and so on. At high speeds, air or lubricating fluids can be expelled from between the meshing teeth at almost sonic speeds and can even become the dominant source of noise. At low speed and load the sound pressure level from a gear increases by about 3 dB for a doubling of load, while at higher speeds and loads the

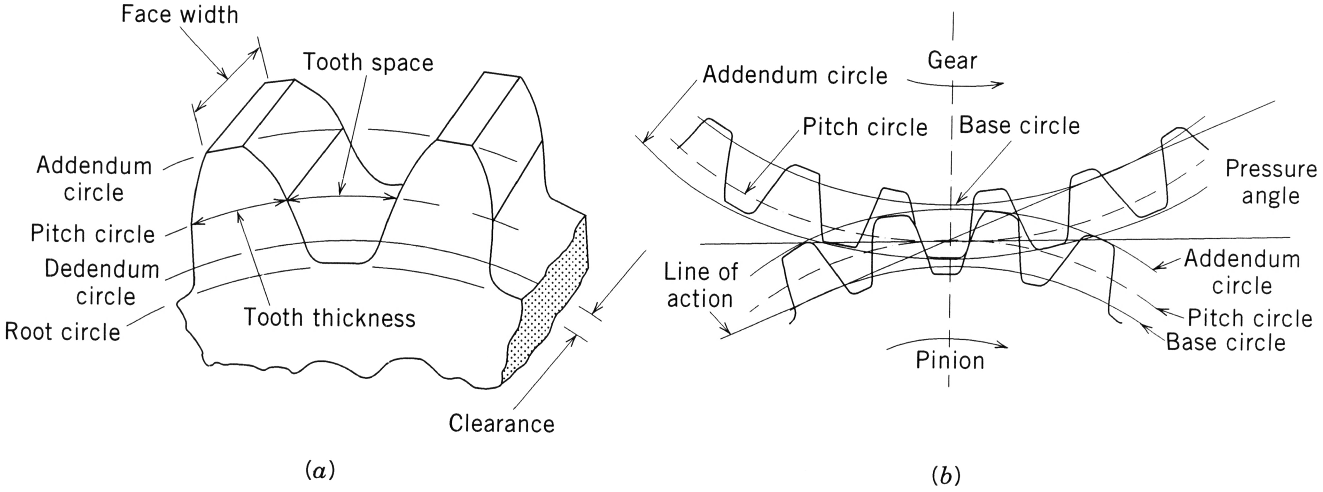

Face width
Tooth space
Addendum circle
Pitch circle
Dedendum circle
Root circle
Tooth thickness
Clearance
(a)
Gear
Addendum circle
Pitch circle
Base circle
Pressure angle
Line of action
Addendum circle
Pitch circle
Base circle
Pinion
(b)

Fig. 23 Terms used with gears: (*a*) involute gear, (*b*) meshing of two parallel-axis spur gears. (Based in part on figures in Refs. 2, 8, and 13.)

sound pressure level increases by about 6 dB for a doubling of speed or load.[18] The noise of a gear set is quite dependent on the quality of manufacture and the tolerances achieved. Thus gear sets can be chosen with very low noise although the cost is high. In many cases where low noise levels are required, it is more cost effective to choose a gear set of moderate cost and to vibration isolate the pinion and gear bearings, apply damping to the gear housing, and, if necessary, completely enclose it. In gears used to transmit only small loads (for exam-

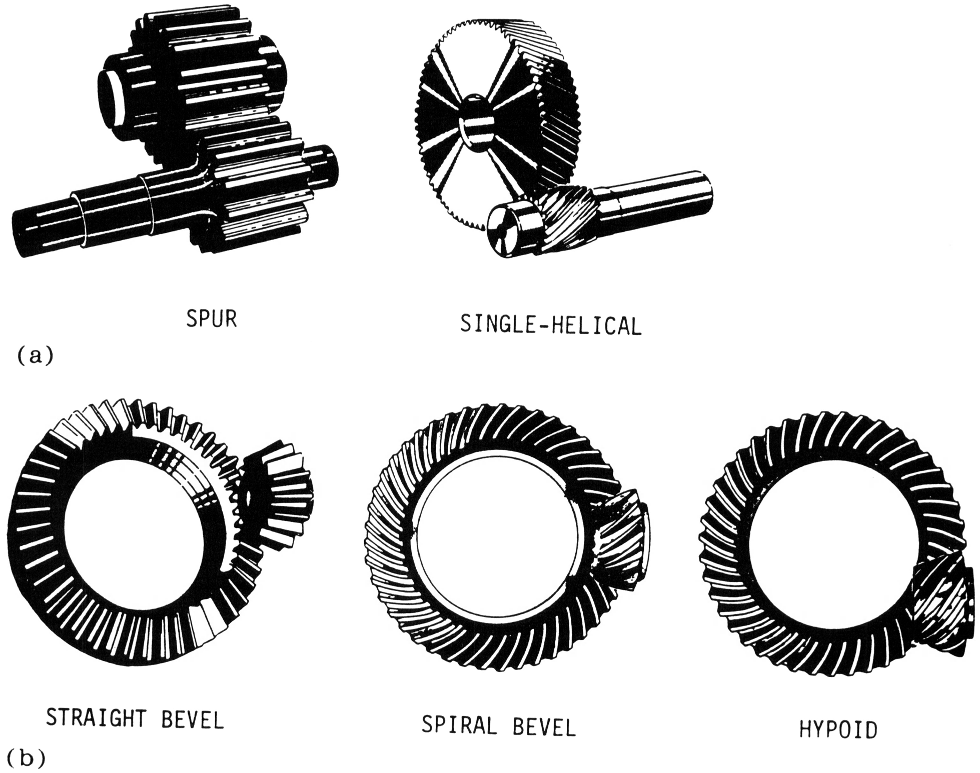

SPUR
SINGLE-HELICAL
(a)
STRAIGHT BEVEL
SPIRAL BEVEL
HYPOID
(b)

Fig. 24 The main types of gear in use: (*a*) parallel axis (spur, helical); (*b*) nonparallel axis (straight bevel, spiral bevel, and hypoid). (Reprinted with permission from R. G. Schegel, R. J. King, and H. R. Mull, "How to Reduce Gear Noise," *Machine Design*, pp. 134–142, 2/27/64, a Penton Publication.)

ple, those in electric clocks), very low noise levels can be achieved by the use of soft gears (which reduce the gear tooth force pulses), and other means.[19] Manufacturing deficiencies can result in variations in pitch and profile from tooth to tooth and eccentricity of the gear wheel, which causes increased noise and vibration.[2,8,13] There would, however, be some noise generated even if the gears were without any imperfections. The frequency of this noise (and vibration) would only occur at the gear meshing frequency f_m and its harmonics:

$$f_m = \frac{N_p n_p}{60}, \quad \text{Hz},$$

where N_p is the number of pinion teeth and n_p is the pinion speed in rpm.

5.3 Fans and Blowers

Fans and blowers are used in appliances, in buildings, in air distribution systems for heating and cooling, and in industry for a variety of purposes.[4] There are two main types of fan design: axial and centrifugal (see Fig. 25). The first three centrifugal fan types: airfoil, backward-curved, and radial are mostly used in industrial applications. The airfoil fan is the most efficient, but it is only suitable for clean-air industrial applications because dust and other particles can adhere to the fan blades and cause malfunctioning. The forward-curved fan is least efficient and usually made of lightweight low-cost materials. The radial fan is the noisiest and least efficient, but is useful for dirty or corrosive flows. Axial fans have the disadvantage that the discharged air is rotating, unless downstream stators are installed; they are mainly used in low- or medium-pressure air conditioning systems. The vaneaxial fan is the most efficient axial type and can operate at high pressure. The propeller type is the least expensive of the axial type. It has the lowest efficiency and is normally limited to low-pressure, high-volume flow applications.

The primary purpose of a fan is to move a required volume flow rate of air against a given back pressure with maximum efficiency and low first cost. There may be additional reqirements such as high resistance to abrasion, ability to transport dusty air, and ease of manufacture or maintenance and repair.

Fan noise has pure-tone and broadband frequency components. Although the mathematical theory of fan noise is well developed, it is beyond the scope of this chapter, and physical explanations are presented instead. Each time a blade passes a point in space, an impulsive force fluctuation is experienced. If a fan has n blades and the rotational speed is N rpm, then the number of impulses experienced per second f_B is

$$f_B = \frac{nN}{60}, \quad \text{Hz}. \tag{23}$$

The frequency f_B is known as the *blade-passing frequency* or often the *blade frequency* for short. Since the time history of the impulsive force at the point will not be completely sinusoidal, harmonics will appear. The strength of the harmonics will be expected to be affected by upstream or downstream solid-body flow obstructions. Fan sound power level data are normally provided by manufacturers in eight one-third-octave bands from 63 to 8000 Hz. If noise is of major concern and the fan has been properly installed, it will probably be necessary to install intake and discharge sound attenuators (with flexible vibration breaks at attachment points to duct systems, if present.)

It is also possible to provide further noise attenuation by the use of ducting, elbows, and plenum chambers lined with sound absorbing material. Care must be taken that significant noise is not generated in this ductwork.

5.4 Electrical Equipment

Examples of electrical equipment that cause noise and vibration include motors, generators and alternators,[12,16] transformers, relays, solenoids, and circuit breakers. Electric motors are used widely in appliances, vehicles, and industry in a variety of types and sizes. They may be commutated, synchronous, or induction types. Electrical energy is converted into mechanical energy, and in the process some heat is produced. Fans are often provided to remove the heat and are the main sources of noise in electric motors. Because of the requirement for most motors that they should operate in either direction, they are usually provided with axial fans, the noisiest type. The sources of noise and vibration in electrical equipment are mostly aerodynamic, mechanical, and electromagnetic in nature. They are summarized in Table 4.

5.5 Control Valves

In industrial plants the aerodynamic noise generated by control valves, regulators and orifices is a major noise source in piping systems. In addition hydrodynamic noise (which is mostly due to cavitation) is of some importance. However in most situations, the aerodynamic noise is dominant.[20] The main noise generating mechanisms include: turbulent mixing, turbulence interaction with boundary, shock waves, interaction of turbulence with shocks, cavity resonances, flow separation, vortex shedding and mechanical vibration.[20] The main noise generating processes can be divided into two

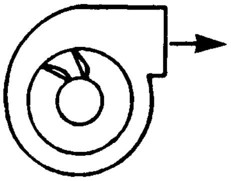
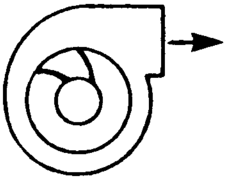
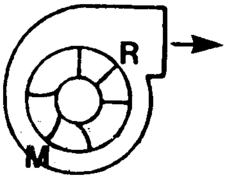
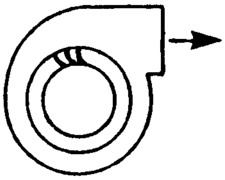
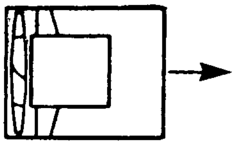
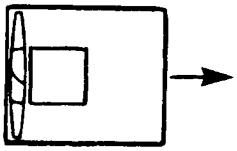
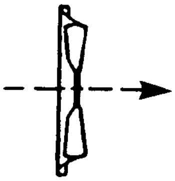
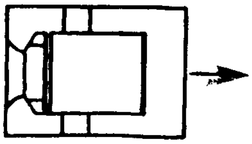
	Fan Type	Description	Design details
CENTRIFUGAL TYPE FANS	Airfoil	Uses 10 to 16 airfoil shape blades; has highest efficiency of all centrifugal types; used where horsepower savings will be important; can be used on low, medium, and high-pressure systems	
	Backward inclined backward curved	Uses 10 to 16 blades; used for similar applications as airfoil fan; gas flow should be clean but need not be as clean as gas flow with airfoil fans	
	Industrial (radial)	Uses 6 to 10 blades of either radial (R) or modified radial (M) type; has lowest efficiency of centrifugal types; used mainly in industrial applications where gas is hot and dirty	
	Forward curved	Uses 24 to 64 blades; construction is usually low-cost and lightweight; efficiency is lower than airfoil or backward curved fans; usually operates at lowest speed and is smallest of centrifugal types; used in low-pressure heating, ventilating, and air-conditioning (HVAC) systems	
AXIAL TYPE FANS	Vaneaxial	Uses 3 to 16 blades; high efficiency airfoil blades may be fixed or adjustable; used in low-medium-, and high-pressure systems in HVAC and other industrial applications	
	Tubeaxial	Uses 4 to 8 blades; usually more efficient than propeller type below and can operate at a higher pressure; does not use guide vanes as does vaneaxial type; Used in low- and medium-pressure systems in HVAC and other applications	
	Propeller type	Uses 2 to 8 blades usually in a circular ring or orifice plate; efficiency low, this type usually limited to low-pressure, high-volume flow applications such as exhaust or attic fans	
	Tubular centrifugal	The fan wheel is usually similar in design to that of the airfoil or backward curved type used in centrifugal fans; since the air is discharged radially and must turn 90° in the guide vane section, its efficiency is lower than similar centrifugal fans; normally used in low-pressure HVAC return air systems	

Fig. 25 The main types of fan with descriptions of their use and design.

TABLE 4 The Main Sources of Noise in Electric Motors

Mechanical	Excessive bearing clearance Nonround bearings Rotor unbalance Rotor eccentricity Crooked shaft Brush and brush holder vibration Misalignment Loose laminations
Electromagnetic	Magnetostriction Torque pulsations Air gap eccentricity Air gap permeance variation Disymmetry
Aerodynamic	Fan blade-passing frequency Turbulence Noise due to airflow path restrictions

regimes: *subsonic*, consisting of turbulence-boundary interaction noise, and *supersonic*, consisting of broadband shock noise. Although the flow is confined by the piping, the turbulent mixing noise is similar to the noise of a free jet discussed in Chapters 27 and 28 and is essentially quadrupole in nature. The shocks are caused by incorrectly expanded flow after the valve which is operating above the critical pressure ratio. The shock noise has two main parts: screech and broadband noise. The screech is discrete in nature and is caused by a feedback mechanism and is not often encountered with valves and regulators. The broadband shock noise is common, however, and has been shown to be mostly independent of flow velocity and to be a function of pressure ratio. Valve noise can be reduced by a number of approaches including design of valves with multiple streams, arranging for the pressure drop to occur through several stages, and using absorptive silencers and thicker pipe walls and pipe lagging. The multiple stream valve designs and absorptive silencers can become plugged by solid particles and moisture can be a problem in the silencer material so that effectiveness is lost. Several authors have given good reviews of valve and piping system noise and the reader is referred to these for more detailed discussion.[20–22]

5.6 Internal Combustion Engines

The internal combustion engine is a major source of noise in transportation and industrial use. The intake and exhaust noise can be effectively silenced. However, the noise radiated by engine surfaces is more difficult to control. In gasoline engines a fuel air mixture is compressed to about one-eighth to one-tenth of its original volume and ignited by a spark plug. In diesel engines air is compressed to about one-sixteenth to one-twentieth of its original volume and liquid fuel is injected in the form of a spray, then spontaneous ignition and combustion occurs. Because the rate of pressure rise is initally more abrupt with a diesel engine than with a gasoline engine, diesel engines tend to be noisier than gasoline engines. The noise of diesel engines has consequently received more attention from both manufacturers and researchers. The noise of engines can be divided into two main parts: combustion noise and mechanical noise. The combustion noise is caused mostly by the rapid pressure noise caused by ignition, and the mechanical noise is caused by a number of mechanisms with perhaps piston slap being one of the most important, particularly in diesel engines. The noise radiated from the engine structure has been found to be almost independent of load, although dependent on cylinder volume and even more dependent on engine speed. Measurements of engine noise over a wide range of cylinder capacities have suggested that engine noise increases by about 17.5 dB(A) for a tenfold increase in cylinder capacity.[10] Engine noise has been found to increase at an even greater rate with speed than with capacity (at least at twice the rate) with about 35 dB(A) for a tenfold increase in speed. Engine noise can be reduced by attention to details of construction. In particular stiffer engine structures have been shown to reduce radiated noise. Partial add-on shields and complete enclosures have been demonstrated to reduce diesel engine noise of the order of 3 to 10 dB(A). Priede has given a good review of internal combustion engine noise with an emphasis on diesel engine noise (see Chapter 19 in Ref. 10).

6 NOISE SOURCE IDENTIFICATION

6.1 Introduction

In most machinery noise control problems, a knowledge of the dominant noise sources in order of importance is very desirable so that suitable modifications can be made. In a complicated machine, such information is often difficult to obtain and many noise reduction attempts are made based on inadequate data so that frequently expensive or inefficient noise reduction methods are employed. Machine noise can also be used to diagnose increased wear. The methods used to identify noise sources will depend on the particular problem and the time and resources (personnel, instrumentation, and money) and expertise available and on the accuracy required. In most noise source identification problems, it is usually the best practice to use more than one method to ensure greater confidence in the identification procedure. In Section 6 of this chapter both well-tried and more novel methods are reviewed.

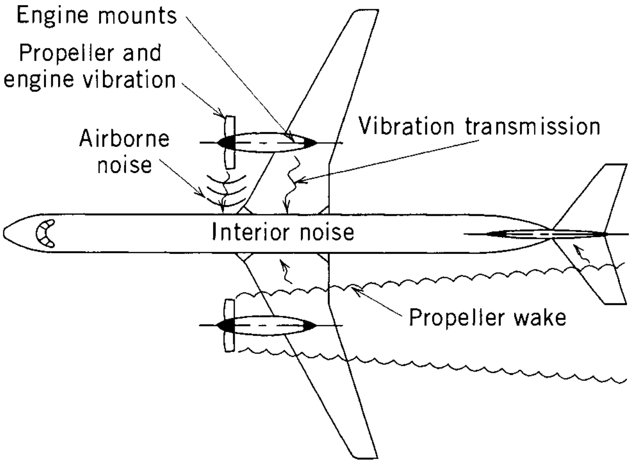

Engine mounts
Propeller and engine vibration
Airborne noise
Vibration transmission
Interior noise
Propeller wake

Fig. 26 Sources and paths of airborne and structure-borne noise resulting in interior noise in an airplane cabin. (Based in part on Fig. 37 in Ref. 33.)

6.2 Source–Path–Receiver

As discussed in Section 2 of this chapter, noise energy must flow from a *source* through one or more *paths* to a *receiver* (usually the human ear). Figure 1 shows the simplest model of a *source–path–receiver* system. Noise sources may be *mechanical* in nature (impacts, out-of-balance forces in machines, vibration of structural parts) or *aerodynamic* in nature (pulsating flow, flow–structure interaction, jet noise, turbulence). Noise will flow through a variety of airborne and structure-borne paths to the receiver. Figure 26 shows an example for a propeller-driven airplane. This airplane situation can be idealized as the much more complicated source–path–receiver system shown in Fig. 27. In some cases it is not easy to separate neatly the sources from the paths, and sources and paths must be considered in conjunction. In other cases the distinction between the source(s) and path(s) is not completely clear. However, the source–path–receiver model is a useful concept that is widely used. In Section 6 of this chapter we will mainly concern ourselves with noise sources. However, we shall also consider the fact that *cutting* or *blocking* one or more noise *paths* often gives invaluable information about the noise sources.

6.3 Clues to Noise Sources

There are various characteristics of a sound field that give indications of the type of noise sources present and their spatial distribution. These include the frequency distribution of the sound pressure level, the directional properties of the sound field, and the variation of the sound pressure level with distance from the sources. A knowledge of the variation in the sound pressure level with time can also be instructive. For more sophisticated measurements it is useful to have some theoretical understanding of the propagation of sound from the idealized models, monopole, dipole, quadrupole, line source, piston, and the generation of sound caused by mechanical systems: mechanical impacts, and vibrating beam and plates. See Chapters 1, 9, and 11. Such theoretical under-

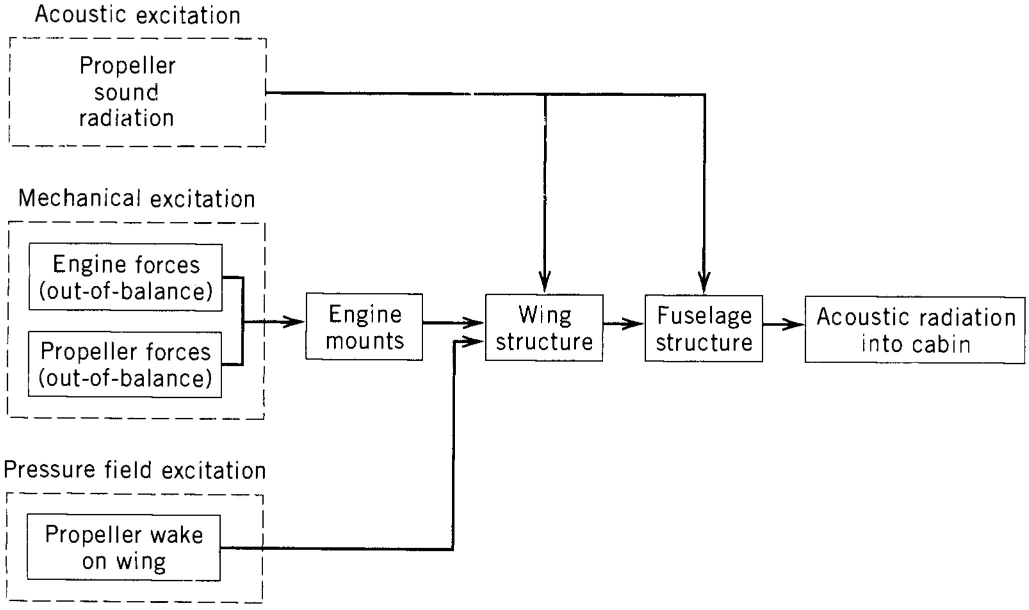

Acoustic excitation
Propeller sound radiation
Mechanical excitation
Engine forces (out-of-balance)
Propeller forces (out-of-balance)
Engine mounts
Wing structure
Fuselage structure
Acoustic radiation into cabin
Pressure field excitation
Propeller wake on wing

Fig. 27 Source–path–receiver system showing airborne and structure-borne paths for a twin-engine propeller driven airplane. (From Ref. 33.)

standing not only guides us in deciding which measurements to make, but also aids in interpreting the results. In fact as the measurements made become more sophisticated, increasing care must be taken in their interpretation. There is a danger with sophisticated measurements of collecting large amounts of data and then either being unable to interpret them or reaching incorrect conclusions. With the advent of minicomputers and other sophisticated signal processing equipment, a knowledge of random process and signal processing theory is very important. See Chapters 100 and 101.

7 CLASSICAL METHODS OF IDENTIFYING NOISE SOURCES

Some of the elementary or classical methods of identifying noise sources have been reviewed in detail in the literature[23] and do not require a profound theoretical understanding of sound propagation. They will only be briefly reviewed here.

Subjective assessment of noise is very useful because the human ear and brain can distinguish between different sounds more precisely than the most sophisticated measurement system. With practice, an operator can tell by its sound if a machine is malfunctioning; a noise consultant can accurately estimate the blade passage frequency of a fan. In identifying sources, we should always listen to a machine first. In order to localize the sources and cut out extraneous noise, stethoscopes can be used, or a microphone–amplifier–earphone system. Such a system also sometimes allows one to position the microphone near a source in an inaccessible or dangerous place where the human ear cannot reach.

Selective operation of a complicated machine is often very useful. As long as the operation of a machine is not severely changed by disconnecting some of the parts sequentially, such a procedure can indicate the probable contribution of these parts to the total machine noise when all parts are operating simultaneously. For instance, engine noise can be measured with and without the cooling fan connected, and this can then give an estimate of the fan noise. Note, however, that if the fan noise is lower than the engine noise in level, the accuracy of the estimate will be poor. The fan noise estimate can be checked by driving the fan separately by a "quieted" electric motor or other device. We should be aware that in some cases disconnecting some parts may alter the operation of the other parts of the machine and give misleading results. An example is an engine driving a hydraulic transmission under load. Disconnecting the transmission will give a measure of the engine-structure radiated noise, provided the inlet and exhaust noise are piped out of the measuring room and the cooling fan is disconnected. But note that the engine is now unloaded, and its noise may be different from the loaded case. Also an unfueled engine may be driven by a "quieted" electric motor so that the so-called engine "mechanical" noise is measured and the "combustion" noise is excluded. However, in this procedure the mechanical forces in the engine will be different from the normal fueled running situation, and we cannot say that this procedure will give the "true" "mechanical" noise.

Selective wrapping or enclosure of different parts of a machine is a useful procedure often used. If the machine is completely enclosed with a tight-fitting sealed enclosure and the parts of the enclosure are sequentially removed exposing different machine surfaces, then, in principle, the noise from these different surfaces can be measured. This method has the advantage that it is not normally necessary to stop any part of the machine, and thus the machine operation is unchanged. However, some minor changes may occur if any damping of the machine surfaces occurs or because sound will now diffract differently around the enclosure.

Measurements of the sound power radiated from different surfaces of a diesel engine using this selective wrapping approach are discussed in Refs. 24 and 25. An array of 30 microphone positions on a spherical surface surrounding the engine was used for both the lead-wrapping and bare-engine sound measurements.

Overall and one-third octave sound pressure level data were gathered. These data were then converted to sound intensity by assuming (1) spherical propagation, (2) no reflections, and (3) that the intensity is equal to the mean square pressure divided by the air impedance ρc. Then the sound intensity data were integrated numerically over the surface of the spherical measuring surface enclosing the engine to obtain the sound power.

The lead was wrapped on the engine so that later a particular surface, for example the oil pan, could be exposed. The one-third octave band results for the fully wrapped engine compared with the results of the bare engine at the 1500 engine speed, 542-N-m torque operating condition are shown in Fig. 28*a*. This comparison shows that the lead wrapping is ineffective at frequencies at and below the 250 Hz one-third octave band.

Eight individual parts of the engine were chosen for noise source identification and ranking purposes using the measurement technique with the spherical microphone array and the engine lead-wrapped. The measurements were made by selectively exposing each of the eight parts, one at a time, while the other seven parts were encased in the 0.8 mm foam-backed lead. Sound power level measurements were then made for each of the eight parts for three separate steady-state operating conditions. One-third-octave band comparisons with the fully wrapped engine are shown in Figs. 28*b* and 28*c* for two of the eight parts at the 1500 rpm operating condition. These plots show that the sound power measurements using lead wrapping are not always accurate for the weaker parts until the frequency exceeds 1000 Hz. A plot of the noise source ranking using overall sound power levels for the eight parts under investigation for the 315-Hz to 10-kHz one-third-octave bands is given in Fig. 29.

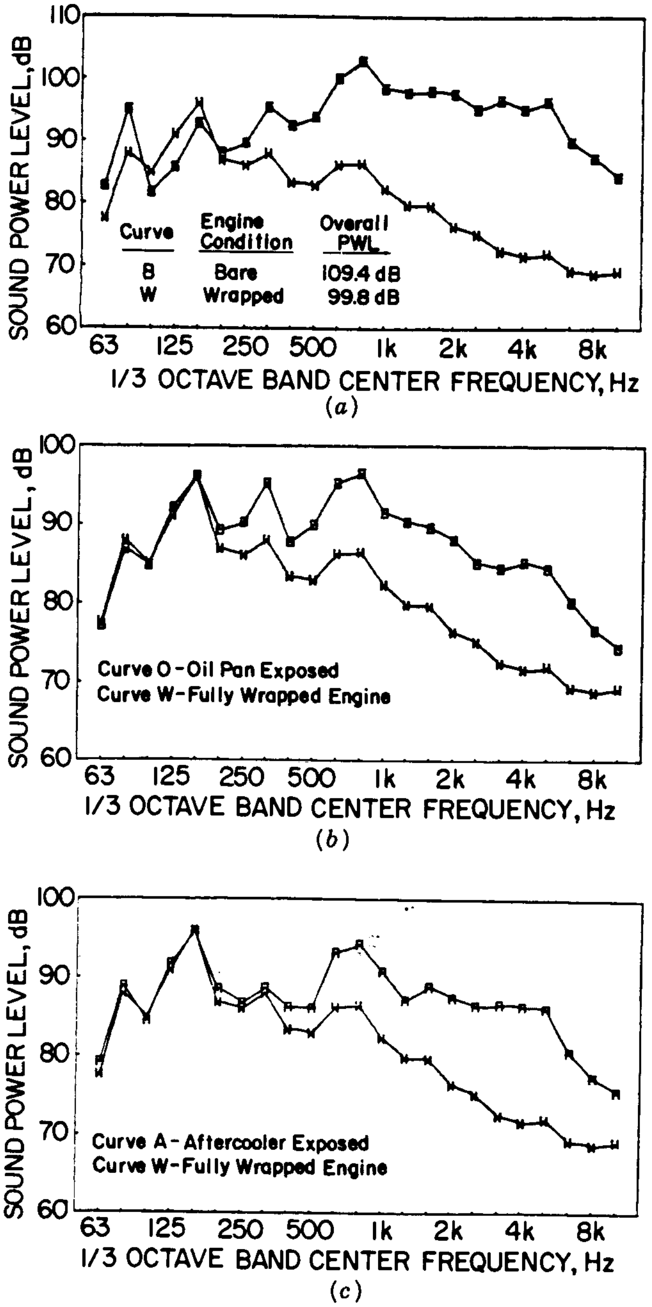


Fig. 28 Sound power level of engine and parts at 1500 rpm and 542-N-m torque: (*a*) bare and fully wrapped engine, (b) oil pan and fully wrapped engine, (c) aftercooler and fully-wrapped engine. (From Ref. 24.)

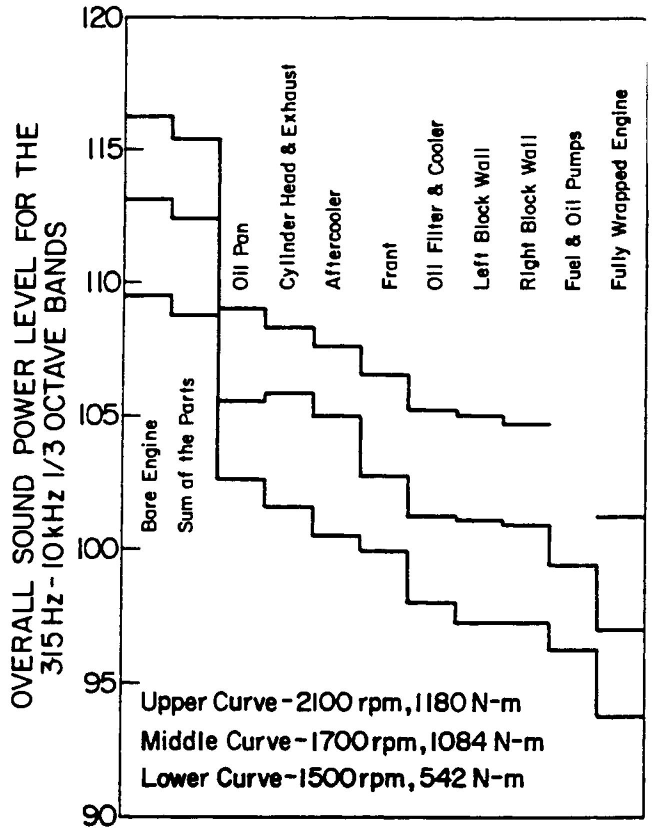


Fig. 29 Noise source ranking in terms of sound power levels from lead-wrapping far-field spherical array method for the 315 Hz to 10 kHz one-third-octave bands. (From Ref. 24.)

Similar selective wrapping techniques have been applied to a complete truck (Fig. 30). Exhaust noise was suppressed with an oversize muffler, the engine was wrapped, and even the wheel bearings were covered! The results of this investigation[26] are shown in Figures 31*a* and 31*b*. Although the selective wrapping technique is very instructive, it is time consuming and this has led to the search for other techniques that do not require complete enclosure, as described later.

8 FREQUENCY ANALYSIS

If the noise (or vibration) of either the complete machine or of some part (obtained by selective operation or

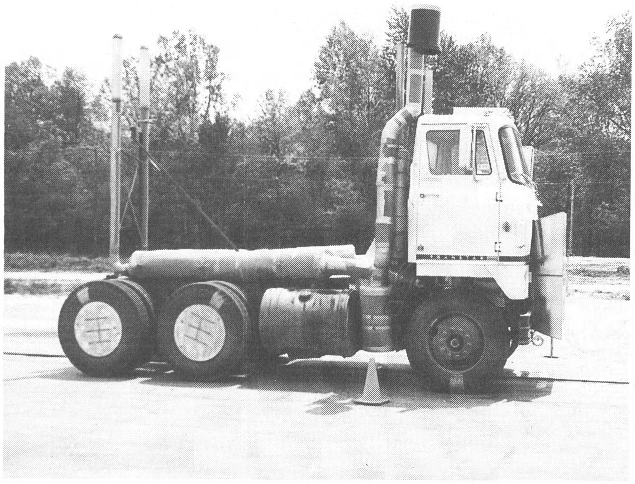


Fig. 30 Selective wrapping and oversize muffler used on a diesel engine truck.

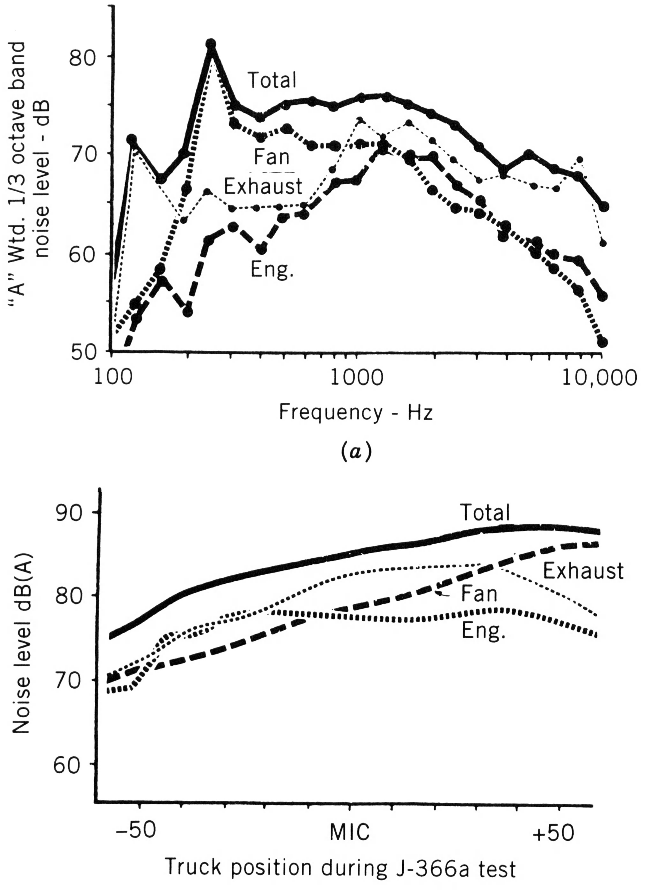


Fig. 31 A-weighted sound levels obtained from selective wrapping and selective operation techniques applied to truck in Fig. 30: (*a*) one-third-octave analysis, and (*b*) A-weighted overall sound level contributions of truck sources as a function of truck drive by position (in feet). (From Ref. 26.)

sequential wrapping) is measured, then various analyses may be made of the data. The most common analysis performed is *frequency analysis*. The reason probably is because one of the most important properties of the ear is performing frequency analysis of sounds. Analog analyzers (filters) have been available for many years. More recently real-time analyzers and fast Fourier transform (FFT) analog/digital minicomputer analyzers have become available.

The analog and real-time analyzers may be either of the parallel or the swept filter type, and these filters may be either of the constant percentage or the constant bandwidth type. With analog filters, we should note that the standard error in the measurements is inversely proportional to the square root of the bandwidth and the averaging time. Thus care should be taken to use large averaging times with narrow-band measurements. With real-time and FFT analyzers different problems arise including aliasing and leakage errors.

The real-time and FFT analyzers speed up the frequency analysis of the data, but because of the analysis errors that are possible (particularly with the FFT) great care should be taken. With FFT analyzers anti-aliasing filters can be used to prevent frequency folding, and if reciprocating machine noise is analyzed, the analysis period must comprise the repetition period (or multiples) or incorrect frequencies are diagnosed.

Assuming the frequency analysis has been made correctly then it can be an important tool in identifying noise sources. With reciprocating and rotating machinery, pure tones are found that depend on the speed of rotation N in rpm, and geometrical properties of the systems. As discussed in Section 5 of this chapter, several frequencies can be identified with rolling element bearings[13,15,27] including the fundamental frequency $N/60$ Hz, the inner and outer race flaw defect frequencies and the ball flaw defect, and the ball resonance and the race resonance frequencies. With gears, noise occurs at the tooth contact frequency $tN/60$ Hz and integer multiples (where t is the number of gear teeth). Often the overtone frequencies are significant. In fan noise, the fundamental blade-passing frequency $nN/60$ (where n is the number of fan blades) and integer multiples are important noise sources. Further machine noise examples could be quoted. Since these frequencies just mentioned are mainly proportional to machine rotational speed, a change of speed or load will usually immediately indicate whether a frequency component is related to a rotational source or not. A good example here is diesel engine noise. Figure 32*a* shows the time history and power spectrum of the cylinder pressure of a six-cylinder direct injection engine,[28] measured by Seybert. Ripples in the time history are thought to be caused by a resonance effect in the gas above the piston producing a broad peak in the frequency spectrum in the region of 3 to 5000 Hz. See Fig. 32*b*. That this resonance frequency *is* related to a gas resonance and not to the rotational speed can clearly be demonstrated since if the speed is changed the frequency is almost unchanged (for constant load), although if the load is increased (at constant speed) the frequency increases. On further investigation the resonance frequency is found to be proportional to the square root of the absolute temperature of the gas (the gas temperature increases with load) and thus as the load increases, so does the resonance frequency.

Frequency analysis can sometimes be used with advantage to reveal information about noise or vibration sources and/or paths by carefully changing the source or the path in some controlled way. The following discussion gives some examples on an air conditioner and an airplane.

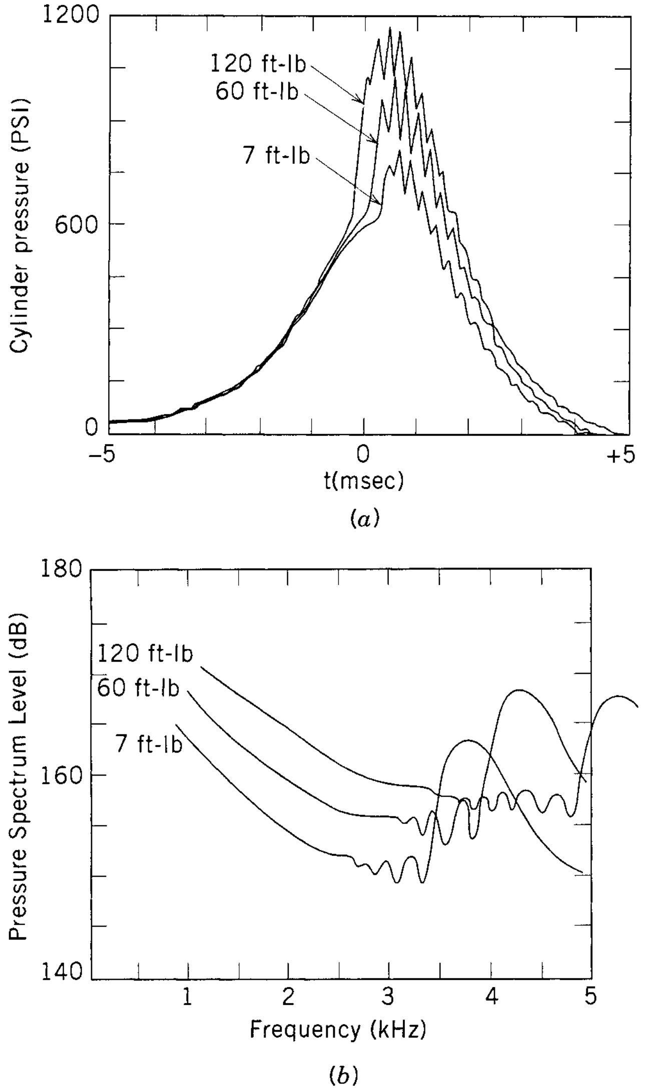


Fig. 32 (*a*) Diesel engine cylinder combustion pressures versus crank angle for three engine load conditions at 2400 rpm, (*b*) combustion pressure spectra at 2400 rpm. (From Ref. 28.)

Change of Excitation Frequency This can give guidance on sources of noise and resonances in a system, as already described with a diesel engine, when the load (and speed) are changed. In some cases, e.g., electrical machines, changing speed is difficult. In the case of the noise from an air conditioner, mechanical resonances in the system can be important.[29] These could be examined by exciting the air conditioner structure at different frequencies with a vibrator. However, this means changing the mechanical system in an unknown way. The approach finally adopted in Ref. 29 was to drive the air conditioner electric motors with an alternating current that could be varied from 50 to 70 Hz from a variable speed DC motor–AC generator set. Since the noise problem was caused by pure-tone magnetic force excitation at twice line (mains) frequency, this was monitored with a microphone at the same time as the motor acceleration. Figure 33 shows the results. This experiment showed that the motor mounts were too stiff. Reducing the stiffness eliminated the pure-tone noise problem. Hague[30] has shown how this technique can be improved further in air-handling systems by using an electric motor in which rotational speed, line frequency, and torque magnitude (both steady and fluctuating) can be varied independently.

In the case just discussed with the air conditioner, it was shown that the transmission path was through the motor mount (which was too stiff). In cases where airborne and structure-borne noise are both known to be problems and the source is known (an engine), the dominant path can sometimes be determined quantitatively by cutting one or more of the paths or by modifying the paths in a known way. Examples include engine noise

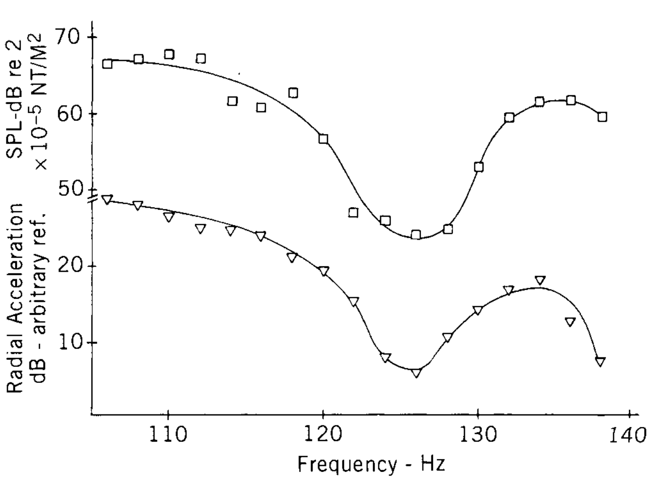


Fig. 33 Sound pressure level and motor acceleration of condenser fan. (From Ref. 29.)

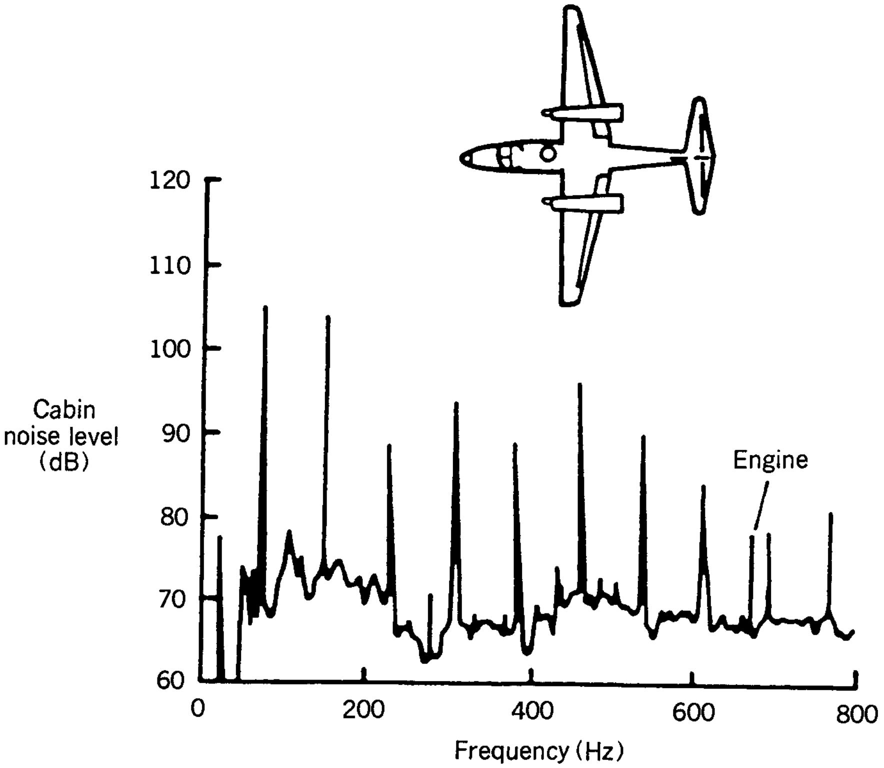


Fig. 36 Spectrum of cabin noise in flight of twin engine airplane (from Ref. 34).

addition of lead sheets or mass-loaded vinyl sheets of 5 to 10 kg/m^2. Figure 35, for example, shows areas of the cabin walls of airplanes that were covered to reduce the airborne transmission.[33]

The spectrum of the noise in the cabin of a light twin-engine airplane is shown in Figure 36. Discrete frequency tones associated with the blade passage frequency and harmonics are evident with the first two being dominant. The broadband noise at about 70 dB is associated with boundary layer noise. The tone at about 670 Hz can be related to the engine turbine speed and suggests the presence of engine-generated structure-borne noise.[34]

9 OTHER CONVENTIONAL METHODS OF NOISE SOURCE IDENTIFICATION

Acoustic ducts have been successfully used by Thien[35] as an alternative to the selective wrapping approach (see Fig. 37). One end of the duct is carefully attached to the part of the machine under examination by a "sound-proof" elastic connection. Thien claims accurate repeatable results from this method. However, one should note various difficulties. Sealing the duct to the machine may be difficult. The surroundings should be made as anechoic as possible to prevent contamination from any reverberant field. Theoretically the duct may alter the acoustic radiation from the machine part examined except for the frequency region much above coincidence. Fortunately this frequency region is the one normally encountered with heavy machinery, and since in this region radiation is mainly perpendicular to the surface, it may be little affected in practice.

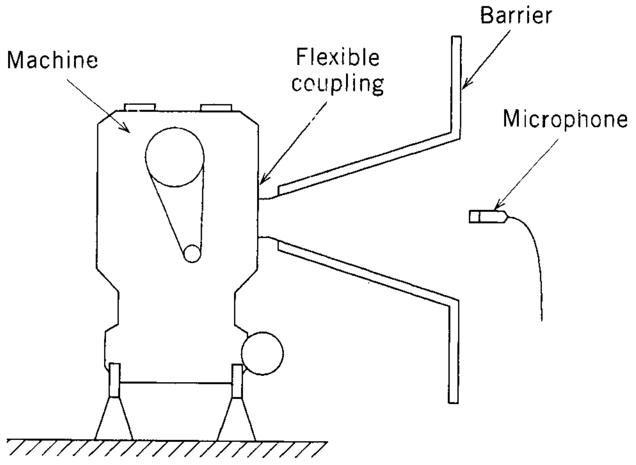


Fig. 37 Acoustical duct used to identify sources.

Acoustic ducts or guides have also been used to investigate paths of noise. Figure 38 shows an example where

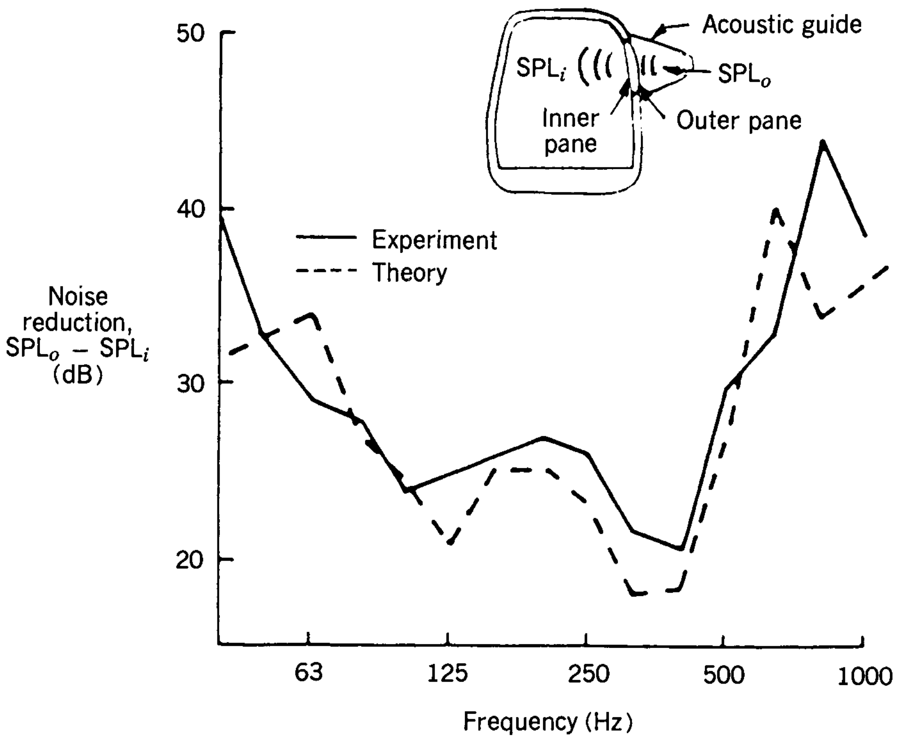


Fig. 38 Investigation of airborne noise path through cabin wall of light airplane using acoustic guide or duct. (From Ref. 36.)

airborne noise was directed onto the cabin wall of a light airplane. The noise reduction through the wall of the airplane was measured and compared with theory. The agreement between the experimental results and the modal theory was excellent at frequencies below about 250 Hz.[36]

Near-field measurements are very often used in an attempt to identify major noise sources on machines. This approach must be used with extreme care. If a microphone is placed in the acoustic near field where kr is small (k is the wavenumber = frequency $f \div$ wavespeed c, and r is the source to microphone distance), then the sound intensity is not proportional to sound pressure squared, and this approach gives misleading results. This is particularly true at low frequency. The acoustic near field is reactive: The sound pressure is almost completely out-of-phase with the acoustic particle velocity.

Thus, this near-field approach is unsuitable for use on relatively "small" machines such as engines. However, for "large" machines such as vehicles with several major sources including an engine cooling fan, exhaust, or inlet, the near-field approach has proved useful.[37] In this case where the machine is large, of characteristic dimension l (length or width), it is possible to position the microphone so that it is in both the near geometric field, where r/l is small, and the far acoustic field, where kr is reasonably large (except at low frequency). In this case the microphone can be placed relatively close to each major noise source. Now the sound intensity is proportional to the sound pressure, and the well-known inverse square law applies. Besides the acoustic near-field effect already mentioned there are two other potential problems with this approach: (i) source directivity and the need to use more than one microphone to describe a large noise source and (ii) contamination of microphone signals placed near individual sources by sound from other stronger sources. Unfortunately, *contamination* cannot be reduced by placing the microphones closer to the sources because then the acoustic near field is increased. However, *contamination* from other strong sources can be allowed for by empirical correction using the inverse square law and the distance to the contaminating source.

Despite the various potential problems with the near-field method, it has been used quite successfully to identify the major noise sources on a large diesel engine truck.[37] By placing microphones near each major noise source (Fig. 39) and then extrapolating to a position 50 ft. (14.5 m) from the truck (Fig. 40) good agreement could be obtained with separate *selective operation* and *selective wrapping* drive-by measurements (Fig. 41). Averaging over five positions near each source and correcting for contaminating sources gave quite close agreement (Fig. 42).

Surface velocity measurements have been used by

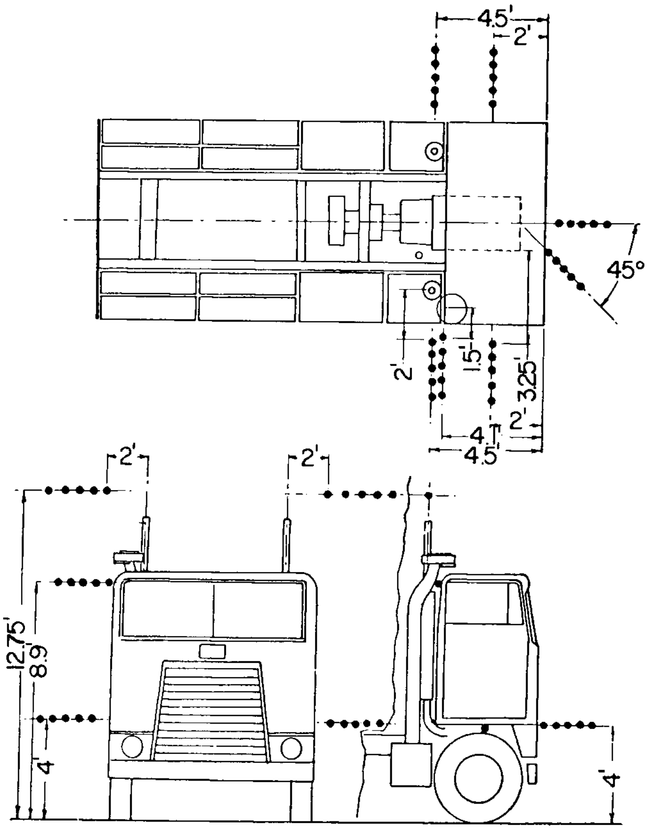


Fig. 39 Microphone positions used in near-field measurements on a truck. (From Ref. 37.)

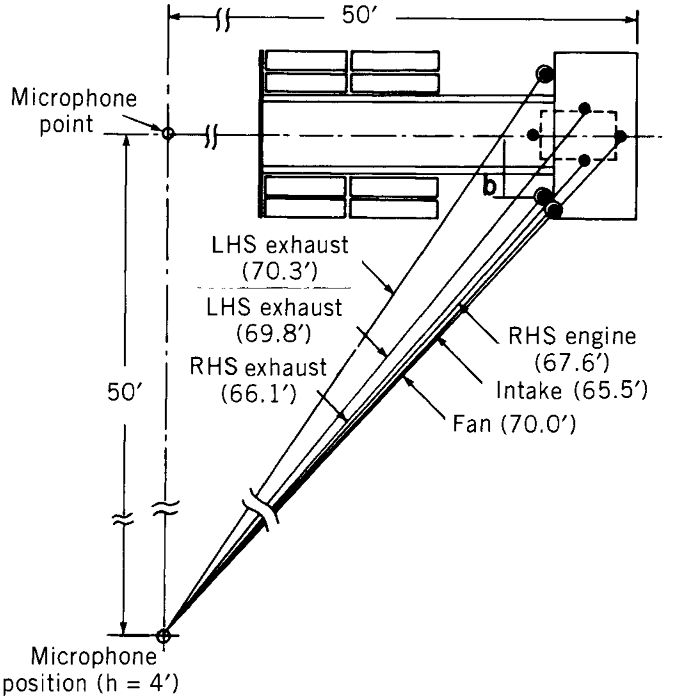


Fig. 40 The position (in feet) of truck (in Fig. 39) in drive-by test when truck engine reached its governed speed of 2100 rpm in sixth gear. Extrapolations to soft measurement location shown. (Note: 1′ = 1 ft., LHS = left-hand side, RHS = right-hand side). (From Ref. 37.)

several investigators to try to determine dominant sources of noise on engines and other machines. The sound power W_{rad} radiated by a vibrating surface of area S is given by

$$W_{\mathrm{rad}} = \rho c S \langle v^2 \rangle \sigma_{\mathrm{rad}}, \tag{24}$$

where ρc is the air characteristic impedance, $\langle v^2 \rangle$ the space-average of the mean-square normal surface velocity, and σ_{rad} the radiation efficiency. Chan and Anderton[38] have used this approach. By measuring the sound power and the space-averaged mean-square velocity on several diesel engines they calculated the radiation efficiency σ_{rad}. They concluded that above 400 Hz for most diesel engines that σ_{rad} was approximately 1. There was a scatter of ±6 dB in their results, although this was less for individual engines. Since σ_{rad} is difficult to calculate theoretically for structures of complicated geometry such as a diesel engine, the assumption that $\sigma_{\mathrm{rad}} = 1$ can thus give an approximate idea of the sound power radiated by each component.

Mapping of the sound level around a machine has been shown[39] to be a useful simple procedure to identify

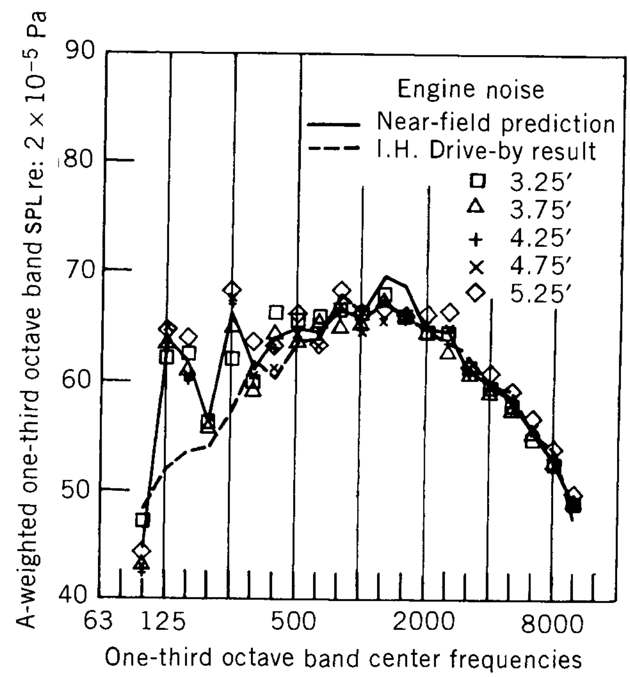


Fig. 41 Near field measurements (shown in Figs. 39 and 40) extrapolated to 50 ft. for comparisons with drive-by noise source identification procedure based on SAE J366a test. (From Ref. 37.)

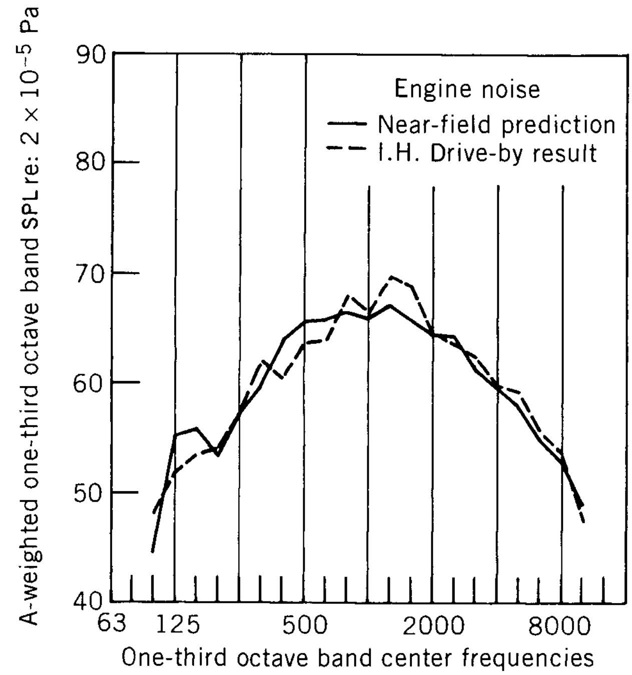


Fig. 42 Engine noise at 50 ft. predicted from near field measurements (see Fig. 41) after corrections are made to eliminate the contribution from the exhaust system noise, compared with the drive-by result. (From Ref. 37.)

major noise sources. For sources radiating hemispherically, it is easy to show that the floor area within a certain level contour is proportional to the sound power output of the machine. This method is graphic and probably more useful to identify the noisiest machines in a workshop than the noisiest parts on the same machine because of the contamination problems between different sources in the latter case. Reverberation in a workshop will affect and distort the lower sound level contours.

10 USE OF CORRELATION AND COHERENCE AND TECHNIQUES TO IDENTIFY NOISE SOURCES AND PATHS

Correlation and coherence techniques in noise control date back over 40 years to work of Goff[40] who first used the correlation technique to identify noise sources. Although correlation has often been used in other applications, it has not been widely used in noise source identification. One exception is the work of Kumar and Srivastava, who reported some success with this technique in identifying noise sources on diesel engines.[41] Since the ear acts as a frequency analyzer, the corresponding approach in the frequency domain (coherence) instead of the time domain is usually preferable. Crocker and Hamilton have reviewed the use of the coherence technique in modeling diesel engine noise.[42] Such a coherence model can also be used, in principle, to give some information about noise sources on an engine. Figure 43 shows an idealized model of a multiple-input, single-output system. It can be shown[42] that the auto spectrum the output noise S_{yy} is given by

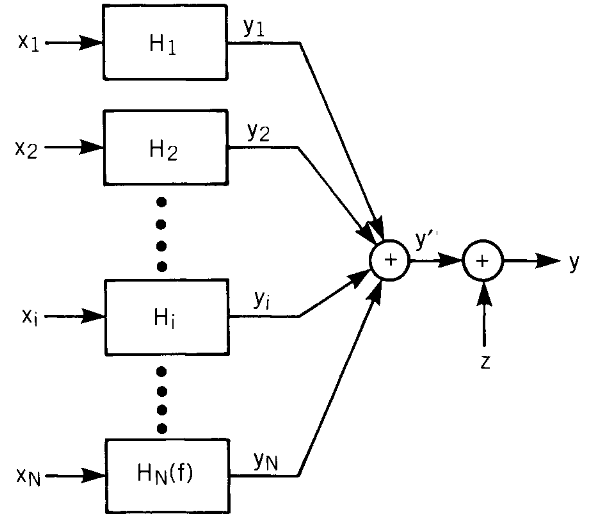


Fig. 43 Multiple-input, single-output system with uncorrected noise $z(t)$ in output.

$$S_{yy} = \sum_{i=l}^{N} \sum_{i=l}^{N} S_{ij} H_i H_j + S_{zz} \tag{25}$$

where N is the number of inputs, the S_{ij} are cross-spectral densities between inputs, H_i and H_j are frequency responses, and S_{zz} is the autospectral density of any uncorrelated noise z present at the output. Note that the S and H terms are frequency dependent.

If there are N inputs, then there will be N equations:

$$S_{iy} = \sum_{i=l}^{N} H_j S_{ij} \qquad i = 1, 2, 3, \ldots, N \tag{26}$$

The frequency responses H_1, H_2, $H_3, \ldots, H_N$ can be found by solving the set of N equations (26).

Several researchers have used the coherence approach to model diesel engine noise.[42–45] This approach appears to give useful information on a naturally aspirated diesel engine, which is combustion-noise dominated and where the N inputs are $x_1, \ldots, x_N$, the cylinder pressures measured by pressure transducers in each cylinder. In this case H_i, the frequency response (transfer function) between the ith cylinder and the far-field microphone can be calculated. Provided proper frequency averaging is performed, the difficulty of high coherence between the cylinder pressures can be overcome because the phasing between the cylinder pressures is exactly known.[46] In this case the quantity $S_{ii}|H_i|^2$ may be regarded as the far-field output noise contributed by the ith cylinder. This may be useful noise source information for the engine designer.

Hayes et al.[47] in further research have extended this coherence approach to try to separate combustion noise from piston-impact noise in a running diesel engine.

Wang and Crocker showed that the multiple coherence approach could be used successfully to separate the noise from sources in an idealized experiment such as one involving three loudspeakers, even if the source signals were quite coherent.[48,49] However, when a similar procedure was used on the more complicated case of a truck which was modeled as a six-input system (fan, engine, exhaust(s), inlet, transmission) the method gave disappointing results and the simpler near-field method appeared better.[37,48] The partial coherence approach was also used in the idealized experiment and the truck experiments. It is believed that contamination between input signals and other computational difficulties may have been responsible for the failure of the coherence method of identifying truck-noise sources.

11 USE OF SOUND INTENSITY FOR NOISE SOURCE AND PATH IDENTIFICATION

Most studies[25,40] have shown that sound intensity measurements are quicker and more accurate than the selective wrapping approach in identifying and measuring the strength of noise sources.

The sound intensity I is the net rate of flow of acoustic energy per unit area. The intensity I_r, in the r direction is

$$I_r = \langle pu_r \rangle, \tag{27}$$

where p is the instantaneous sound pressure, u_r is the acoustic particle velocity in the r direction, and the angle brackets denote a time average. The sound power W radiated by a source can be obtained by integrating the component of the intensity I_n normal to any surface enclosing the source:

$$W = \int I_n \, dS \tag{28}$$

The sound pressure p and particle velocity are normally determined with two closely spaced microphones (1 and 2) and the surface S is a surface enclosing the source.

Chung first used this technique with the surface of a diesel engine subdivided into $N = 98$ radiating areas.[50] The power was determined from each area in about two minutes each. The engine was source-identified in less than a day, which involved much less time than the selective wrapping technique. Since the intensity measurements were made only 20 mm above the engine surfaces, an anechoic or semi-anechoic room was unnecessary. Many other researchers have used this sound intensity approach to identify noise sources since then as is discussed in more detail in Chapter 156.

REFERENCES

1. R. H. Bolt and K. U. Ingard, "System Considerations in Noise Control Problems, in C. M. Harris (Ed.), *Handbook of Noise Control*, McGraw-Hill, New York, 1957, Chapter 22.
2. R. H. Lyon, *Machinery Noise and Diagnostics*, Butterworths, Boston, 1987.
3. D. A. Bies and C. H. Hansen, *Engineering Noise Control*, 2nd ed., Chapman & Hall, London, 1996.
4. M. J. Crocker and F. M. Kessler, *Noise and Noise Control*, Volume II, CRC Press, Boca Raton, 1982, Chapter 1.
5. I. Sharland, *Woods Practical Guide to Noise Control*, Woods of Colchester Limited, printed by Waterlow and Sons Ltd, London, 1972.
6. R. S. Jackson, "The Performance of Acoustic Hoods at Low Frequencies," *Acustica*, Vol. 12, 1962, p. 139.
7. R. S. Jackson, "Some Aspects of the Performance of Acoustic Hoods," *J. Sound Vib.*, Vol. 3, No. 1, 1966, p. 82.
8. L. H. Bell and D. H. Bell, *Industrial Noise Control*, 2nd ed., Marcel Dekker, New York, 1994.
9. C. M. Harris (Ed.), *Handbook of Acoustical Measurements and Noise Control*, 3rd Edition, McGraw-Hill, New York, 1991.
10. L. L. Beranek and I. L. Ver (Eds.) *Noise and Vibration Control Engineering*, Wiley, New York, 1992.
11. S. Skaistis, *Noise Control of Hydraulic Machinery*, Marcel Dekker, New York, 1988.
12. P. L. Timar (Ed.), *Noise and Vibration of Electrical Machines*, Elsevier, Amsterdam, 1989.
13. J. E. Shahan and G. Kamperman, "Machine Element Noise," in *Handbook of Industrial Noise*, Industrial Press, 1976, Chapter 8.
14. M. P. Norton, *Fundamentals of Noise and Vibration Analysis for Engineers*, Cambridge University Press, Cambridge, 1989.
15. P. Eschmann, L. Hasbargen, and K. Weigand, *Ball and Roller Bearings: Theory, Design, and Application*, Wiley, Chichester, 1985.
16. W. B. Rowe, *Hydrostatic and Hybrid Bearings Designs*, Butterworths, London, 1983.

17. D. D. Fuller, *Theory and Practice of Lubrication for Engineers*, 2nd ed. Wiley, Chichester, 1984.
18. L. D. Mitchell, "Gear Noise: The Purchaser's and the Manufacturer's Views. Noise and Vibration Control Engineering." *Proceedings of the Purdue Noise Control Conference*, M. J. Crocker (Ed.), Purdue University, July 1971, pp. 95–106.
19. W. M. Viebrock and M. J. Crocker, "Noise Reduction of a Consumer Electric Clock," *Sound and Vibration*, Vol. 7, No. 3, 1973, pp. 22–26.
20. K. M. Ng, "Control Valve Noise," *ISA Trans.*, Vol. 33, 1994, pp. 275–286.
21. G. Reethof, "Turbulence-Generated Noise in Pipe Flow," *Ann. Rev. Fluid Mech.*, 1978, pp. 333–367.
22. W. F. Blake, *Mechanics of Fluid-Induced Sound and Vibration*, Academic Press, New York, 1986.
23. C. E. Ebbing and T. H. Hodgson, "Diagnostic Tests for Locating Noise Sources," *Noise Control Eng.*, Vol. 3, 1974, pp. 30–46.
24. M. C. McGary and M. J. Crocker, "Surface Intensity Measurements on a Diesel Engine," *Noise Control Eng.*, Vol. 16, No. 1, 1981, pp. 26–36.
25. T. E. Reinhart and M. J. Crocker, "Source Identification on a Diesel Engine Using Acoustic Intensity Measurements," *Noise Control Eng.*, Vol. 18, No. 3, 1982, pp. 84–92.
26. R. L. Staadt, "Truck Noise Control, in Reduction of Machinery Noise," (Rev. Ed.), Short Course Procs., Purdue University, December 1975, pp. 158–190.
27. R. A. Collacott, "The Identification of the Source of Machine Noises Contained within a Multiple-Source Environment," *Appl. Acoustics*, Vol. 9, No. 3, 1976, pp. 225–238.
28. A. F. Seybert, "Estimation of Frequency Response in Acoustical Systems with Particular Application to Diesel Engine Noise," Ph.D. Thesis, Purdue University, Herrick Laboratories Report HL 76-3, December 1975.
29. A. F. Seybert, M. J. Crocker, et al., "Reducing the Noise of a Residential Air Conditioner," *Noise Control Eng.*, Vol. 1, No. 2, 1973, pp. 79–85.
30. J. M. Hague, "Dynamic Vibration Exciter of Torsional Axial, and Radial Modes from Modified D.C. Motor," Paper No. 962 presented at ASHRAE, Semi-Annual Meeting, Chicago, Illinois, February 13–17, 1977.
31. J. F. Unruh et al, "Engine Induced Structural-Born Noise in a General Aviation Aircraft." NASA CR-159099, 1979 (see also SAE Paper 7900626).
32. M. C. McGary and W. H. Mayes, "A New Measurement Method for Separating Airborne and Structureborne Aircraft Interior Noise," *Noise Control Eng.*, Vol. 20, No. 1, 1983, pp. 21–30.
33. J. S. Mixon and J. F. Wilby, "Interior Noise," in H. H. Hubbard (Ed.), *Aeroacoustics of Flight Vehicles: Theory and Practice*, Volume 2: Noise Control, NASA Reference Publication 1258, August 1991, Chapter 16.
34. J. F. Wilby et al., "In-Flight Acoustic Measurements on a Light Twin-Engined Turbo prop Airplane," NASA CR-178004, 1984.
35. G. E. Thien, "The Use of Specially Designed Covers and Shields to Reduce Diesel Engine Noise," SAE Paper 730244, 1973.
36. R. Vailaitis et al., "Transmission Through Stiffened Panels," *J. Sound Vibr.*, Vol. 70, No. 3, 1980, 413–426.
37. M. J. Crocker and J. W. Sullivan, "Measurement of Truck and Vehicle Noise," 1978, SAE Paper 780387, see also SAE Transactions.
38. C. M. P. Chan and D. Anderton, "Correlation Between Engine Block Surface Vibration and Radiated Noise of In-Line Diesel Engines," *Noise Control Eng.*, Vol. 2, No. 1, p. 16, 1974.
39. P. Francois, "Isolation et Revetments," *Les Carte de Niveaux Sonores*, Jan.–Feb. 1966, pp. 5–17.
40. K. W. Goff, "The Application of Correlation Techniques to Source Acoustical Measurements," *J. Acoust. Soc. Am.*, Vol. 27, No. 2, 1955, pp. 336–346.
41. S. Kumar and N. S. Srivastava, "Investigation of Noise Due to Structural Vibrations Using a Cross-Correlation Technique," *J. Acoust. Soc. Am.*, Vol. 57, No. 4, 1975, pp. 769–772.
42. M. J. Crocker and J. F. Hamilton, "Modeling of Diesel Engine Noise Using Coherence," SAE Paper 790362.
43. J. Y. Chung, M. J. Crocker, and J. F. Hamilton, "Measurement of Frequency Responses and the Multiple Coherence Function of the Noise Generation System of a Diesel Engine," *J. Acoust. Soc. Am.*, Vol. 58, No. 3, 1975, pp. 635–642.
44. A. F. Seybert and M. J. Crocker, "The Use of Coherence Techniques to Predict the Effect of Engine Operating Parameters On Noise," *Trans. ASME, J. Eng. Ind.*, Vol. 97B, 1976, p. 13.
45. A. F. Seybert and M. J. Crocker, "Recent Applications of Coherence Function Techniques in Diagnosis and Prediction of Noise," *Proc. INTER-NOISE 76*, 1976, pp. 7–12.
46. A. F. Seybert and M. J. Crocker, "The Effect of Input Cross-Spectra on the Estimation of Frequency Response Functions in Certain Multiple-Input Systems," *Arch. Acoustics*, Vol. 3, No. 3, 1978, pp. 3–23.
47. P. A. Hayes, A. F. Seybert, and J. F. Hamilton, "A Coherence Model for Piston-Impact Generated Noise," SAE Paper 790274.
48. M. E. Wang, "The Application of Coherence Function Techniques for Noise Source Identification," Ph.D. Thesis, Purdue University, 1978.
49. M. E. Wang and M. J. Crocker, "Recent Applications of Coherence Techniques for Noise Source Identification, *Proc. INTER-NOISE 78*, 1978, pp. 375–382.
50. J. Y. Chung et al., "Application of Acoustic Intensity Measurement to Engine Noise Evaluation," SAE Paper 790502, 1979.

84

ACTIVE NOISE CONTROL

P. A. NELSON AND S. J. ELLIOTT

1 INTRODUCTION

Active techniques for controlling noise have developed rapidly during the 1980s and 1990s, largely because of the developments during this period in modern electronics. Relatively inexpensive digital signal processors have become widely available that enable analog audio frequency signals to be converted into digital form, processed via a digital filter, and then converted back into an analog signal with very little time delay. This electronic capability, together with the wave nature of sound, has made the implementation of active noise control a feasible practical proposition.

Relatively simple single-channel systems are now commercially available and have been used successfully, for example, for the control of low-frequency sound in air conditioning systems. In this case, the sound in the air ducts propagates only as plane waves (below a certain frequency) and destructive interference caused by the field produced by a control loudspeaker leads to successful suppression of the sound field in the duct downstream of the loudspeaker. The possibilities for extending these techniques to other types of acoustic field has stimulated considerable research in recent years, much of it summarized in Ref. 1. It is the physical nature of sound fields that most often dictates the limits to the performance of active techniques, and it is with this in mind that we review below the possibilities for and recent practical applications of active noise control. No attempt has been made here to deal with the suppression of noise through the active control of vibration. This is dealt with comprehensively in Ref. 2 and is also summarized in Chapter 75.

At the time of writing, active control has been successfully applied to the suppression of low-frequency duct-borne broadband sound, low-frequency periodic sound generated by electrical power transformers, aircraft propellers and automobile and turbofan engines, and low-frequency broadband road-generated noise inside road vehicles. However, the design of active control systems remains a specialist activity and is only just reaching the stage of commercialization in a few, mostly high technology, applications. Perhaps the application that has achieved the most widespread commercial production is the "noise-canceling headset," which uses a simple feedback loop to suppress low-frequency sound in the region of the listener's ears. This and other possibilities will be described further below.

2 PRINCIPLES OF ACTIVE CONTROL

A simple single-channel active noise control system is illustrated in Fig. 1. Sound is detected by a microphone and processed via a digital filter (implemented on a special-purpose microprocessor) prior to being input to a loudspeaker that radiates the sound that is intended to destructively interfere with the "primary" unwanted sound. The success of the system is monitored by an "error microphone" placed downstream of the loudspeaker. The characteristics of the digital filter are designed (and often continuously adapted) in order to minimize the time-averaged signal at the error microphone.

In the following description of active techniques we will work from first principles and concentrate initially on the limitations imposed by the physical nature of sound fields. It is of course the principle of superposition that enables active control to be achieved. Thus, any

---

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker
Copyright © 1997 John Wiley & Sons, Inc.

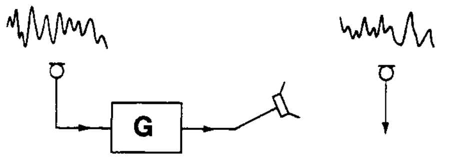


Fig. 1 Single-channel active noise control system showing a detection sensor whose output is passed through a signal processor (G) before being input to a loudspeaker. The success of the system is monitored by an error sensor.

number of sound fields that individually satisfy the governing (linear) wave equation can simply be superposed (in space and time) in order to produce a more desirable sound field. In evaluating the extent to which this can be achieved in practice, it is often useful to work at a single frequency and deal with the spatially dependent complex pressure $p(\mathbf{x})$, which satisfies the homogeneous Helmholtz equation

$$(\nabla^2 + k^2)p(\mathbf{x}) = 0, \qquad (1)$$

where the complex pressure $p(\mathbf{x})$ is the sum of the complex pressures $p_p(\mathbf{x})$ and $p_s(\mathbf{x})$ in the primary and secondary sound fields. The wavenumber k is given by ω/c_0, where ω is the angular frequency of the fluctuation. While it is often convenient to evaluate the potential for active control by working at a single frequency, it should always be recognized that no constraint is necessarily placed on the causal relationship between the secondary and primary fields.[1] Nevertheless, calculations at a single frequency always reveal the best that can be achieved by active techniques.

3 ACTIVE CONTROL OF FREE-FIELD SOUND

3.1 Control of Sound with Continuous Source Layers

Figure 2 shows a primary-source distribution in an unbounded medium specified by the complex volume velocity per unit volume $q_{p,\mathrm{vol}}(\mathbf{y})$, where the vector $\mathbf{y}$ is used as a position vector in the source coordinate system. The primary complex pressure field $p_p(\mathbf{x})$ satisfies the Kirchhoff–Helmholtz integral equation

$$\int_S [g(\mathbf{x}|\mathbf{y})\, \nabla_y p_p(\mathbf{y}) - p_p(\mathbf{y})\, \nabla_y g(\mathbf{x}|\mathbf{y})] \cdot \mathbf{n}\, dS$$

$$= \begin{cases} p_p(\mathbf{x}), & \mathbf{x} \text{ within V}, \\ 0, & \mathbf{x} \text{ outside V}, \end{cases} \qquad (2)$$

where V is a restricted volume of the medium bounded by the surface S and $g(\mathbf{x}|\mathbf{y})$ is the free-space Green function $e^{-jk|\mathbf{x}-\mathbf{y}|}/4\pi|\mathbf{x}-\mathbf{y}|$. The symbol ∇_y represents the del operator in the $\mathbf{y}$-coordinate system. The terms in the surface integral on the left side of this equation can be interpreted as a surface distribution of monopole and dipole sources.[1] Thus, if one places on the surface S a secondary distribution of monopole and dipole sources whose strengths are respectively determined by $\nabla_y p_p(\mathbf{y})$ and $p_p(\mathbf{y})$, then one can generate a field equal to $-p_p(\mathbf{x})$ inside the volume V while leaving the field outside V unchanged. The possibilities for active control that this solution represents were first pointed out by Jessel[3] and Malyhuzinets.[4] While offering a very general solution to the problem of canceling a sound field over a spatial volume, the need to appropriately activate a continuous monopole/dipole source layer renders this technique extremely difficult to implement in practice.

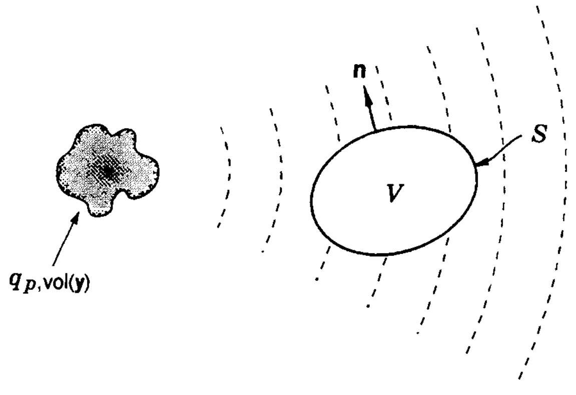
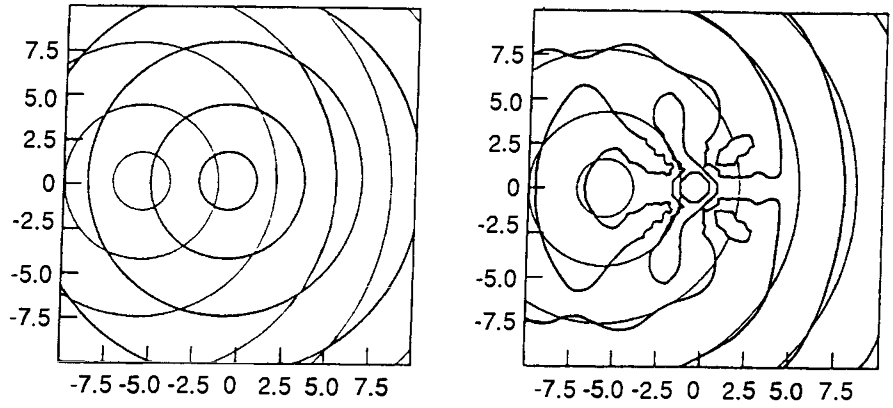


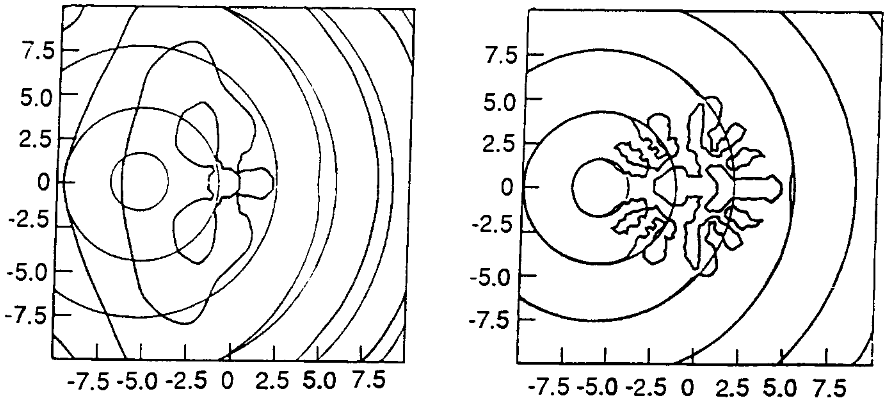
Fig. 2 Continuous monopole–dipole source layer placed on the surface S can be used to produce perfect cancellation of the sound field inside V while leaving the field outside V completely unchanged. Reprinted with kind permission of Academic Press.

3.2 Control of a Compact Primary-Source Distribution with a Series of Point Monopoles

Another possibility for producing complete cancellation of the far-field sound of a primary-source distribution is that suggested by Kempton.[5] For a primary-source distribution that is restricted to a spatial region of dimensions small compared to the acoustic wavelength, the radiated field can be represented by that of a series of point multipoles located at a given position in or close to the source region. It can be shown[6] that the monopole, dipole, quadrupole, and so on, moments associated with the multipole series are determined exactly by the properties of the primary-source distribution. It follows, therefore, that with the multipole moments appropriately determined, a series of multipole sources of opposite



Number of multipoles = 9 Number of multipoles = 9



Number of multipoles = 15 Number of multipoles = 15

Fig. 3 Contours of constant phase of the fields due to a point monopole $(-\varepsilon, 0, 0)$ and a series of multipoles at the origin. The above results were computed by Kempton[5] for a value of $k\varepsilon = 5$. Reprinted with kind permission of Academic Press.

strength when placed at a position close to the primary source will produce cancellation of the far field. The results of some of Kempton's computer simulations[5] are shown in Fig. 3. Again, the difficulty in realizing a practical approximation to a multipole series of such complexity precludes the practical application of the technique.

3.3 Active Control with Discrete Monopole Secondary Sources

A more realistic approach to the active control of free-field sound is to accept that the secondary sources that one can use in practice consist of discrete monopole (loudspeaker) elements. The strengths of these secondary sources are adjusted in order to minimize a quadratic cost function such as the sum of the squared pressures at a discrete number of sensor locations. This approach has been applied by Kido and Onoda,[7] Piraux and Nayroles,[8] and Nelson et al.[9-11] Figure 4 shows L sensors at which the complex pressure produced by the primary source can be specified by the complex vector $\mathbf{p}_p$. The M secondary sources have complex strengths (volume velocities) specified by the complex vector $\mathbf{q}_s$. The total com-

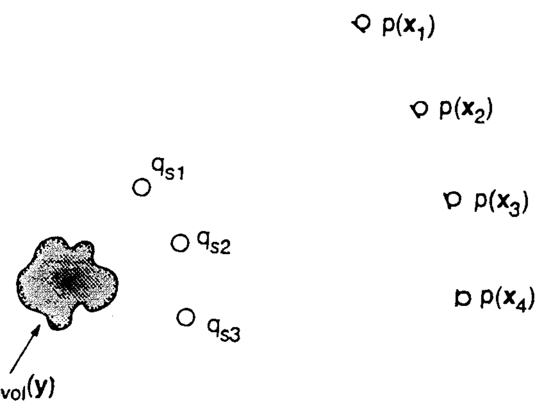


Fig. 4 Active control of sound with discrete secondary sources. The strengths of the secondary sources q_{sm} are adjusted to minimize the sum of the squared pressures $p(\mathbf{x}_l)$.

plex pressure at the sensors is thus given by the vector

$$\mathbf{p} = \mathbf{p}_p + \mathbf{Z}\mathbf{q}_s, \tag{3}$$

where $\mathbf{Z}$ is the $L \times M$ complex transfer impedance matrix that determines the pressure produced at the sensors by the secondary sources. For $M = L$, $\mathbf{p}$ can be made equal to zero by choosing $\mathbf{q}_s = -\mathbf{Z}^{-1}\mathbf{p}_p$ for a nonsingular $\mathbf{Z}$. Another approach, when $L > M$ is to determine $\mathbf{q}_s$, in order to minimize the sum of squared pressures,

$$J_p = \sum_{l=1}^{L} |p(\mathbf{x}_l)|^2 = \mathbf{p}^{\mathrm{H}}\mathbf{p}, \tag{4}$$

where H denotes the Hermitian transpose (complex-conjugate transpose) operation. Substitution of Eq. (3) into Eq. (4) yields

$$J_p = \mathbf{q}_s^{\mathrm{H}}\mathbf{A}\mathbf{q}_s + \mathbf{q}_s^{\mathrm{H}}\mathbf{b} + \mathbf{b}^{\mathrm{H}}\mathbf{q}_s + c, \tag{5}$$

which is a Hermitian quadratic function of the vector $\mathbf{q}_s$ with the matrix $\mathbf{A} = \mathbf{Z}^{\mathrm{H}}\mathbf{Z}$, the complex vector $\mathbf{b} = \mathbf{Z}^{\mathrm{H}}\mathbf{p}_p$, and the scalar $c = \mathbf{p}_p^{\mathrm{H}}\mathbf{p}_p$. If the matrix $\mathbf{A}$ is positive definite, it can be shown[9] that this function has a unique minimum J_{p0} corresponding to an optimal vector $\mathbf{q}_{s0}$. These are given by

$$J_{p0} = c - \mathbf{b}^{\mathrm{H}}\mathbf{A}^{-1}\mathbf{b}, \tag{6}$$

$$\mathbf{q}_{s0} = -\mathbf{A}^{-1}\mathbf{b}. \tag{7}$$

These equations provide an unequivocal definition of the best that can be done to reduce the sum of the mean-squared pressures in the sound field. One important result that follows from an analysis of this type is the minimum total power output to the sound field when the field of a point monopole primary source is controlled with a point monopole secondary source, that is, the minimum of the sum of the squared pressures produced at an infinite number of sensors on a surface in the far field of the source pair. It can be shown[9] that no appreciable reductions in power output can be achieved if the sources are separated by a distance that is greater than one-half wavelength. This result serves to emphasize that active techniques are most often successful at low frequencies.

4 ACTIVE CONTROL OF SOUND IN DUCTS

4.1 Active Control of Sound Propagation in an Infinite Rigid Waveguide

The complex pressure in the infinite rigid-walled waveguide illustrated in Fig. 5 can be expressed in terms of the eigenfunctions (mode shape functions) $\psi_n(x_1, x_2)$ of the duct cross section. Thus, for $x_3 \gg y_3$ we may write[12]

$$p(\mathbf{x}) = \sum_{n=0}^{N} \frac{\omega\rho_0\psi_n(x_1,x_2)}{2\kappa_n L_1 L_2} e^{-j\kappa_n y_3} \cdot \int_V q_{\mathrm{vol}}(\mathbf{y}) e^{j\kappa_n y_3} \psi_n(y_1, y_2)\, dV, \tag{8}$$

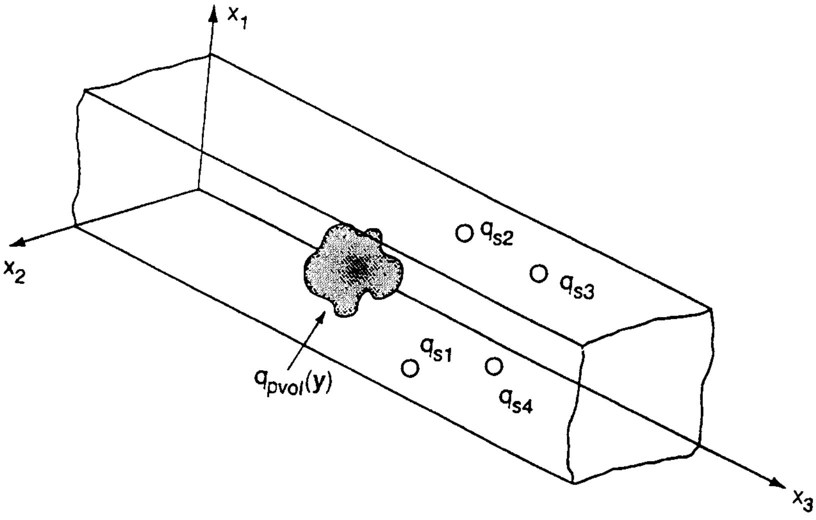


Fig. 5 Active control of sound propagating in a duct using a number of point secondary sources.

where ρ_0 is the density, the wavenumber

$$\kappa_n = \sqrt{k^2 - \left(\frac{n_1\pi}{L_1}\right)^2 - \left(\frac{n_2\pi}{L_2}\right)^2}$$

(the integers n_1 and n_2 characterizing a given mode), and the summation is taken over only the N propagating modes for which κ_n is real. If, as illustrated in Fig. 5, the source distribution is given by some primary-source distribution $q_{p,\mathrm{vol}}(\mathbf{y})$ plus a number M of discrete secondary sources, the source strength distribution may be written as

$$q_{\mathrm{vol}}(\mathbf{y}) = q_{p,\mathrm{vol}}(\mathbf{y}) + \sum_{m=1}^{M} q_{sm}\delta(\mathbf{y} - \mathbf{y}_m), \tag{9}$$

where $\delta(\cdot)$ represents the three-dimensional Dirac delta function. Substitution into Eq. (8) shows that

$$p(\mathbf{x}) = \sum_{n=1}^{N} a_n\psi_n(x_1, x_2) = \mathbf{a}^{\mathrm{T}}\boldsymbol{\psi}, \tag{10}$$

where $\boldsymbol{\psi}$ is a vector of mode shape functions and $\mathbf{a}$ is a vector of complex-mode amplitudes given by

$$\mathbf{a} = \mathbf{a}_p + \mathbf{B}\mathbf{q}_s. \tag{11}$$

The vector $\mathbf{a}_p$ defines the complex-mode amplitudes produced by the primary source only, while the $N \times M$ matrix $\mathbf{B}$ determines the mode amplitudes produced by the secondary sources whose complex strengths are characterized by the vector $\mathbf{q}_s$. In the case $N = M$, with the secondary sources located to ensure that $\mathbf{B}$ is nonsingular, it follows that choosing $\mathbf{q}_s = -\mathbf{B}^{-1}\mathbf{a}_p$ ensures that $\mathbf{a} = 0$ and thus $p(\mathbf{x}) = 0$. Thus, if at a given frequency the number of secondary sources is chosen to equal the number of propagating modes, the sound field well downstream of the secondary sources can in principle be made equal to zero. Successful controller implementations based on this strategy have only recently been reported.[13,14]

4.2 Active Control of Plane-Wave Sound

In ducts excited at frequencies less than the lowest cut-on frequency $c_0\pi/L_1$ (assuming $L_1 > L_2$), only the plane-wave mode will propagate. The active control of such sound fields has been the subject of extensive research (see, e.g., Refs. 15–18). It follows from the above analysis that a single secondary source, when appropriately driven, will be sufficient to ensure zero pressure well downstream of the secondary source. It can be shown[1] that such a strategy ensures that downstream propagating plane waves are reflected back upstream; the secondary source provides a pressure-release boundary condition. A standing-wave field is then produced between the primary and secondary sources. A means of avoiding this was suggested by Swinbanks,[18] who showed that two secondary sources spaced apart along the duct axis would be used to provide perfect absorption of incident plane-wave sound over a limited bandwidth. It can also be shown in the low-frequency limit[1] that the secondary-source strengths required to accomplish this give a monopole–dipole source combination that follows from the general theory presented in Section 3.1.

5 ACTIVE CONTROL OF ENCLOSED SOUND FIELDS

5.1 Control at Low Modal Densities

For acoustic wavelengths that are appreciable compared to the dimensions of the enclosure, it is appropriate to use a modal model of the sound field. The complex pressure can be written as[1,19]

$$p(\mathbf{x}) = \sum_{n=0}^{\infty} \frac{\omega\rho_0 c_0^2 \psi_n(\mathbf{x})}{V[\omega c_0 D_{nn} + j(\omega^2 - \omega_n^2)]} \int_V q_{\mathrm{vol}}(\mathbf{y})\psi_n(\mathbf{y})\, dV, \tag{12}$$

where $\psi_n(\mathbf{x})$ is the mode shape function corresponding to the nth mode, V is the enclosure volume, D_{nn} is a damping term determined by the admittance of the enclosure walls,[1,19] and ω_n is the natural frequency of the nth mode. It is now assumed, as in the case of duct-borne sound described in Section 4.1, that there is some primary source distribution in the enclosure and that the complex pressure can similarly be expressed as the *finite* summation

$$p(\mathbf{x}) = \sum_{n=0}^{N} a_n\psi_n(\mathbf{x}) = \mathbf{a}^{\mathrm{T}}\boldsymbol{\psi}. \tag{13}$$

If it is also assumed that M secondary sources are introduced to control the field, the vector $\mathbf{a}$ can again be expressed in the form

$$\mathbf{a} = \mathbf{a}_p + \mathbf{B}\mathbf{q}_s. \tag{14}$$

At first sight it appears that this problem can also be treated by ensuring that $M = N$ and choosing $\mathbf{q}_s = -\mathbf{B}^{-1}\mathbf{a}_p$ to give $\mathbf{a} = 0$. However, in order to give an adequate representation of the sound field, the number N of modes required in the summation will necessarily be very much greater than the number M of secondary sources. The approach taken[10] is therefore to find the secondary-source strengths that minimize the total acoustic potential energy in the enclosure. This is given by

$$E_p = \frac{1}{4\rho_0 c_0^2} \int_V |p(\mathbf{x})|^2 \, dV = \frac{V}{4\rho_0 c_0^2} \mathbf{a}^{\mathrm{H}}\mathbf{a}, \qquad (15)$$

where the second of these expressions results from the orthogonality property of the acoustic modes. Substitution of Eq. (14) then shows that

$$E_p = \frac{V}{4\rho_0 c_0^2} [\mathbf{q}_s^{\mathrm{H}}\mathbf{B}^{\mathrm{H}}\mathbf{B}\mathbf{q}_s + \mathbf{q}_s^{\mathrm{H}}\mathbf{B}^{\mathrm{H}}\mathbf{a}_p + \mathbf{a}_p^{\mathrm{H}}\mathbf{B}\mathbf{q}_s + \mathbf{a}_p^{\mathrm{H}}\mathbf{a}_p], \qquad (16)$$

which is a quadratic function of $\mathbf{q}_s$ of the type defined by Eq. (5) and has a unique minimum value associated with an optimal vector of secondary-source strengths. Figure 6 shows the results of some computer simulations undertaken by Bullmore et al.[20] Global reductions in energy are produced by a single secondary source placed close to the primary source. Three secondary sources placed appropriately are also shown to produce global reduc-

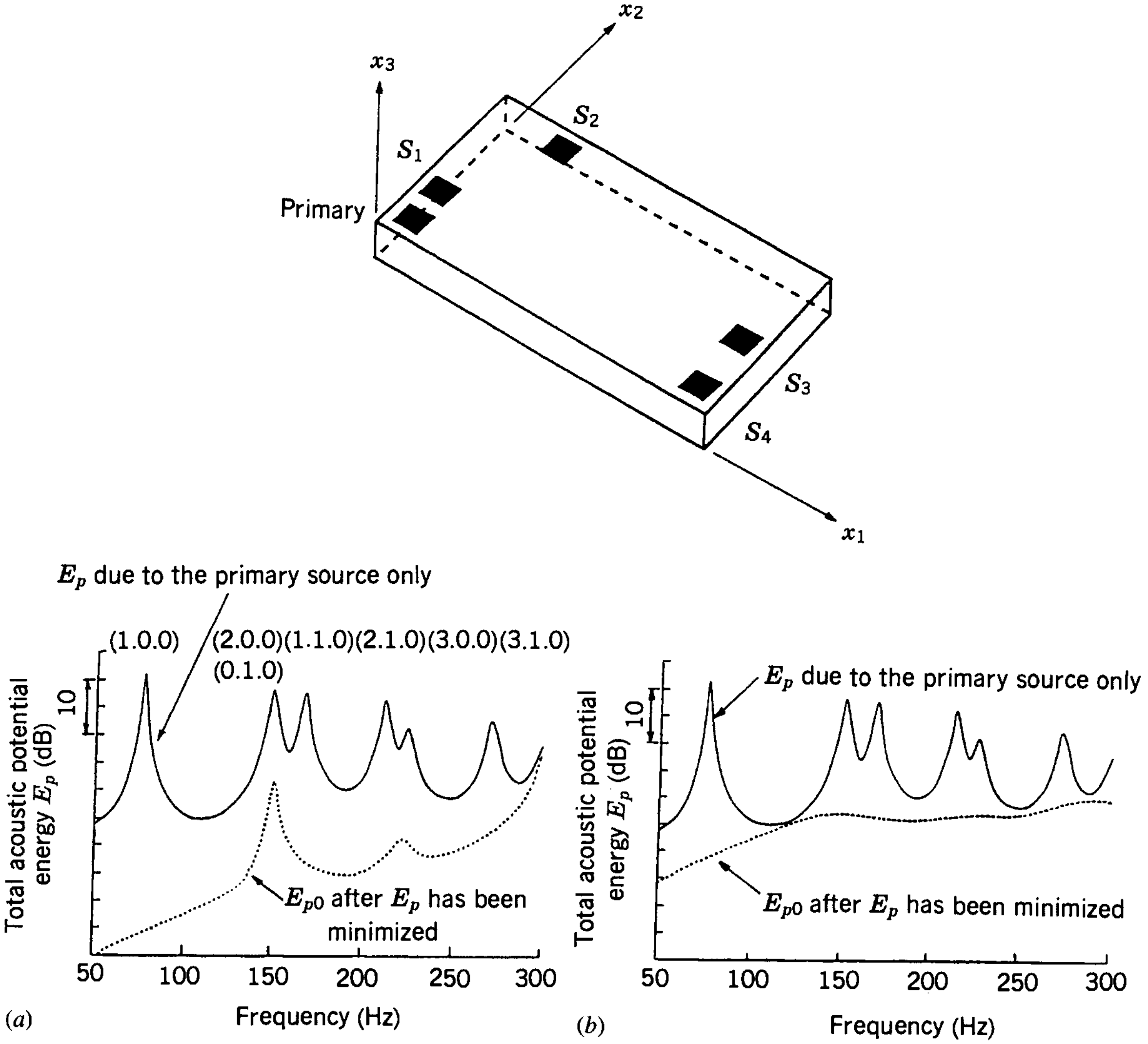


Fig. 6 Shallow rectangular enclosure in which the total acoustic potential energy is minimized by (*a*) source S_1 and (*b*) sources S_2, S_3, and S_4 together. (Results after Ref. 20.) Reprinted with kind permission of Academic Press.

tions in the energy of the enclosure at frequencies close to the natural frequencies of lightly damped modes.

5.2 Control at High Modal Densities

For frequencies of excitation that correspond to acoustic wavelengths that are much smaller than the enclosure dimensions, global control of an enclosed sound field is far more difficult to accomplish. In fact, for frequencies above the Schroeder frequency[21] it can be shown[10] that global reductions in energy in the enclosure can only be produced if secondary sources are separated from the primary source by a distance that is much less than one-half wavelength from the primary source. In practice, in the absence of a highly localized primary source, it may only be possible to produce localized "zones of quiet" around sensors at which the total pressure is minimized by the action of secondary sources. This strategy has been investigated by Elliott et al.,[22] who used computer simulations together with an analysis based on diffuse-field theory to investigate the size of the quiet zone produced around a sensor at which the pressure was driven to zero. The *average* size of the zone in which the sound pressure was reduced to at least 10 dB below that due to the primary source was found to be described by a sphere having a diameter of only one-tenth of an acoustic wavelength.

6 SINGLE-CHANNEL FEEDBACK CONTROL SYSTEMS

One of the earliest attempts to automatically control a sound field was that described by Olson and May,[23] who used a simple high-gain feedback loop to drive the pressure to a low value in the region of a sensor placed close to a loudspeaker. More recently, this approach has been found useful in controlling the sound field in ear defenders[24,25] and in controlling plane-wave sound propagation in ducts.[26,27] Its use in the former case is illustrated in Fig. 7. The electroacoustic frequency response function relating the voltage input to the loudspeaker to the voltage output to the microphone is given by $C(j\omega)$, and the output of the microphone is fed back to the input of the loudspeaker via a filter having the frequency response function $G(j\omega)$. Using the superposition principle applied to the electrical variables involved, the Fourier transform of the microphone output signal can be written as

$$E(\omega) = D(\omega) + G(j\omega)C(j\omega)E(\omega), \qquad (17)$$

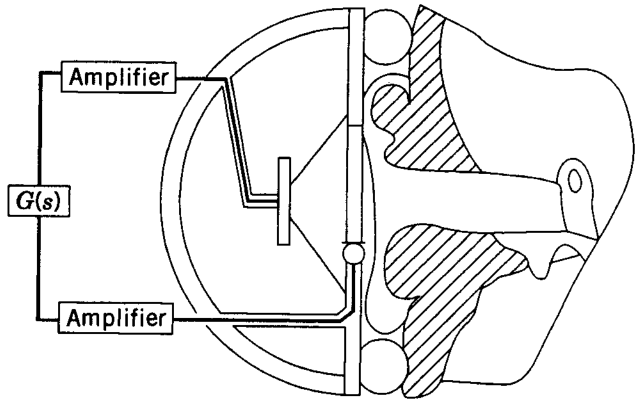


Fig. 7 Feedback control of the sound field at the entrance to the ear canal (after Ref. 25). Reprinted with kind permission of Academic Press.

where $D(\omega)$ is the Fourier transform of the microphone signal produced by the primary field alone. Thus

$$E(\omega) = D(\omega)\left[\frac{1}{1 - G(j\omega)C(j\omega)}\right], \qquad (18)$$

and the spectral densities of $E(\omega)$ and $D(\omega)$ are related by[1]

$$S_{ee}(\omega) = S_{dd}(\omega)\left[\frac{1}{|1 - G(j\omega)C(j\omega)|^2}\right]. \qquad (19)$$

This expression shows that $|1 - G(j\omega)C(j\omega)|^2$ must be made as large as possible if the spectral density of the microphone signal is to be reduced. Thus, in general terms, given a value of $C(j\omega)$, the frequency response function $G(j\omega)$ must be manipulated in order to ensure as high as possible a value of $|G(j\omega)C(j\omega)|$ over as broad a frequency range as possible while still maintaining a stable system. The system stability is governed by the Nyquist stability criterion, which, roughly speaking,[1] requires that the loop gain be less than unity when the loop phase shift is an integer multiple of 2π. Considerable attention must therefore be given to the design of the phase response of the filter $G(j\omega)$, and it is worth noting[1] that many of the successful implementations to date[24,25] have used a minimum-phase filter $G(j\omega)$ with the same number of poles and zeros. Figure 8 shows the results of the system devised by Carme,[25] who used a second-order compensation filter $G(j\omega)$ with considerable success.

7 MULTICHANNEL FEEDFORWARD CONTROL SYSTEMS

7.1 Equivalent Block Diagram

The wave nature of sound leads more naturally to a feedforward approach to control system design, and it is this

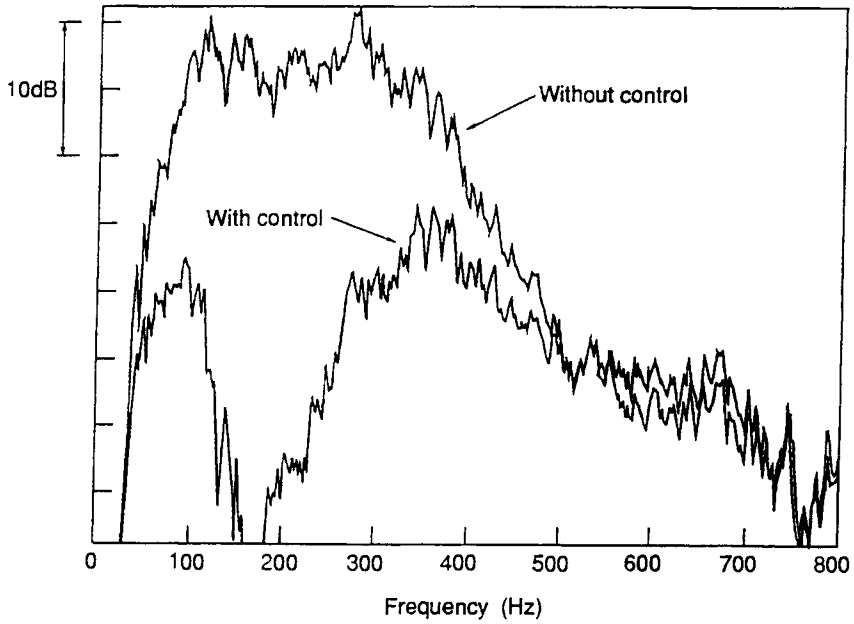


Fig. 8 Reduction in the power spectral density of broadband noise at the entrance to the ear canal produced by the closed-loop system devised by Carme.[25] Reprinted with kind permission of Academic Press.

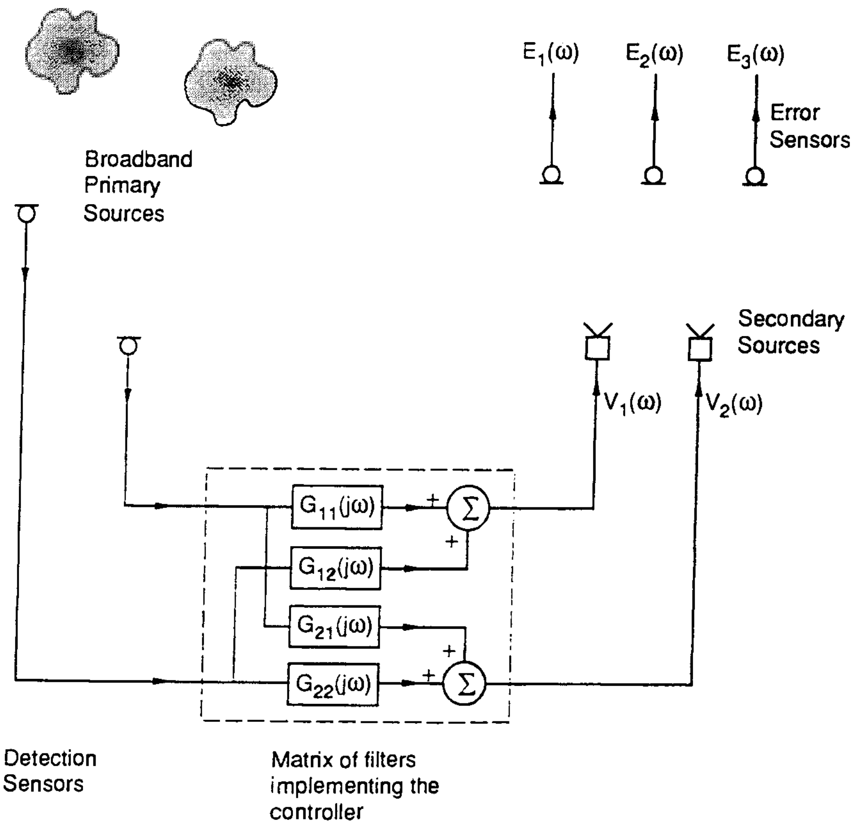

Broadband Primary Sources
$E_1(\omega)$
$E_2(\omega)$
$E_3(\omega)$
Error Sensors
Secondary Sources
$V_1(\omega)$
$V_2(\omega)$
$G_{11}(j\omega)$
$G_{12}(j\omega)$
$G_{21}(j\omega)$
$G_{22}(j\omega)$
+
+
+
+
Σ
Σ
Detection Sensors
Matrix of filters implementing the controller

Fig. 9 The multichannel feedforward control problem. The sound radiated by a number of primary sources is detected by K detection sensors, processed via a matrix of control filters and output via M secondary sources. The success of control is monitored by L error sensors. Note that the output of the secondary sources may also be fed back to the detection sensors. Reprinted with kind permission of Academic Press.

approach that has recently received more attention than feedback control. The general structure of a multichannel feedforward controller is sketched in Fig. 9 and is shown in block diagram form in Fig. 10.

In the block diagram it has been assumed that we can apply the principle of superposition to the electrical variables involved. The vector **e** represents the vector of Fourier transforms of the signals detected by L error sensors. These signals are the sum of the signals **z** produced by the secondary sources and the signals **d** produced by the primary sources. The latter are also assumed to contain measurement noise. The control filters are specified by the matrix **G** of frequency response functions. The M controller output signals, specified by the vector **v** of Fourier transforms, reach the error sensors after passing via the electroacoustic paths specified by the matrix **C** of frequency response functions. They also reach the K detection sensors after passing via the "feedback" electroacoustic paths specified by the matrix **F** of frequency response functions. In the absence of the secondary sources, the detection sensors produce a vector of output signals **u** that contain contributions from the primary sources and measurement noise.

In determining the matrix **G**, it is useful to first determine the matrix **H** of frequency response functions relating **v** to **u**, as depicted in Fig. 10. Once the optimal value

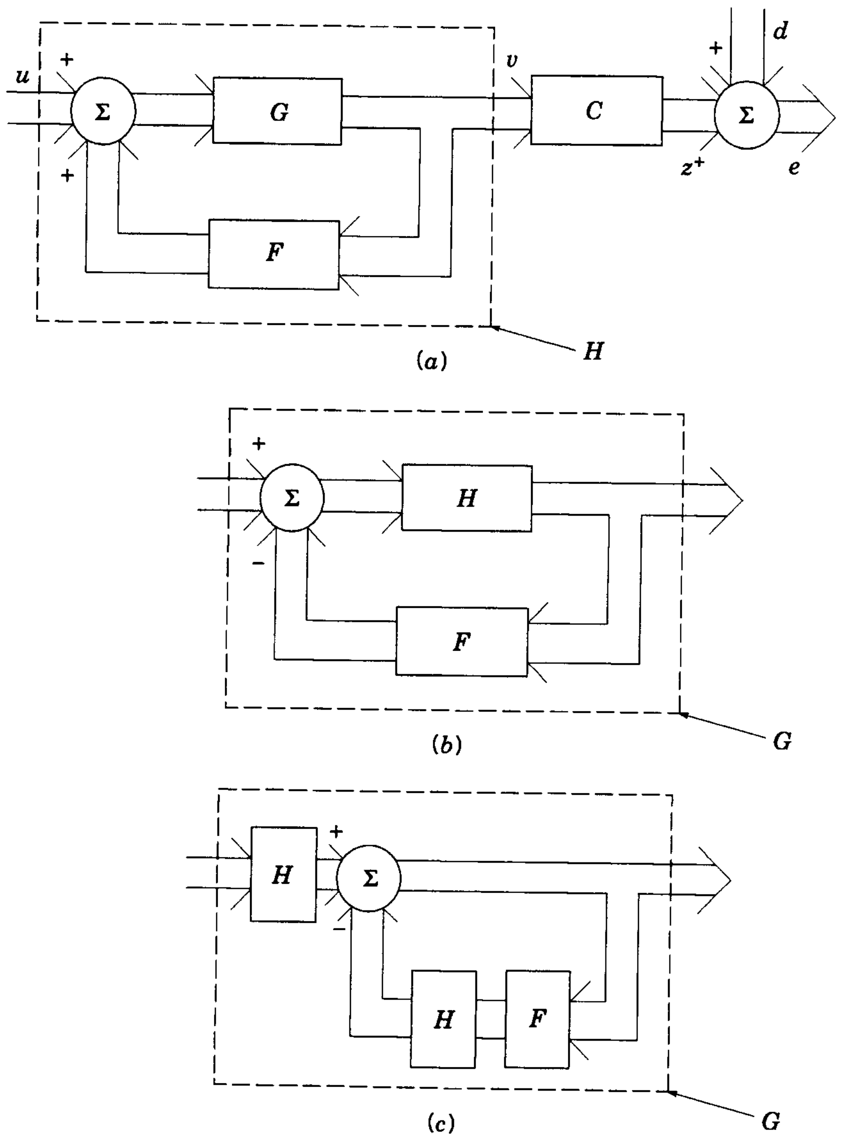


Fig. 10 Block diagram of the multichannel feedforward control problem. The elements of the matrix **H** are first determined. The matrix of control filters **G** can be implemented either using the "feedback cancellation" architecture shown in (*b*) or the "recursive" structure shown in (*c*). Reprinted with kind permission of Academic Press.

of **H** is determined, assuming **F** is known (it can easily be measured), the optimal value of **G** can also in principle be found. It follows from Fig. 10 that

$$\mathbf{v} = \mathbf{G}[\mathbf{u} + \mathbf{F}\mathbf{v}] \tag{20}$$

and therefore that

$$\mathbf{v} = [\mathbf{I} - \mathbf{G}\mathbf{F}]^{-1}\mathbf{G}\mathbf{u}, \tag{21}$$

where **I** is the identity matrix. Thus the matrix **H** can be identified as

$$\mathbf{H} = [\mathbf{I} - \mathbf{G}\mathbf{F}]^{-1}\mathbf{G}. \tag{22}$$

This expression can be rearranged[28] to give two alternative solutions for the matrix **G**. Thus

$$\mathbf{G} = [\mathbf{I} + \mathbf{H}\mathbf{F}]^{-1}\mathbf{H} = \mathbf{H}[\mathbf{I} + \mathbf{F}\mathbf{H}]^{-1}. \tag{23}$$

These expressions can be represented by the filter structures depicted in Fig. 10. The first shows the "feedback cancellation" architecture suggested by the first expression for the matrix **G**, while the second illustrates that the same can be achieved using a "recursive"-type arrangement of filter matrices.

The equivalent block diagram consisting of the cascade of matrices **H** and **C** can also be redrawn by effectively reversing the order of operation of the elements of the matrices. Since the matrices themselves do not commute, we have to first consider the signal produced by the secondary source at the lth sensor. Thus, the Fourier transform of the signal is given by

$$Z_l(\omega) = \sum_{k=1}^{K} \sum_{m=1}^{M} H_{mk}(j\omega) R_{lmk}(\omega), \tag{24}$$

where $R_{lmk}(\omega)$ is the Fourier transform of the *filtered reference signal*, given by

$$R_{lmk}(\omega) = C_{lm}(j\omega) U_k(\omega), \tag{25}$$

which is the signal produced by passing the kth detected signal through the lmth element of the matrix **C**.

7.2 Frequency-Domain Analysis

With the filtered reference signals appropriately defined, it is now possible to undertake an analysis in the frequency domain. This enables the deduction of the best results that can be achieved using a controller that is not constrained to be causal. However, it will be assumed that the signals involved are random with stationary statistical properties, the equivalent results for single frequency sound being readily obtained as a special case. The Fourier transform $Z_l(\omega)$ can be written in the form

$$Z_l(\omega) = \mathbf{h}^{\mathrm{T}}\mathbf{r}_l = \mathbf{r}_l^{\mathrm{T}}\mathbf{h}, \tag{26}$$

where the vectors $\mathbf{r}_l$ and $\mathbf{h}$ are defined by

$$\begin{aligned}\mathbf{r}_l^{\mathrm{T}} = [&R_{l11}(\omega)\cdots R_{lM1}(\omega)|R_{l12}(\omega)\cdots R_{lM2}(\omega)|\\ &\cdots|R_{l1K}(\omega)\cdots R_{lMK}(\omega)],\end{aligned} \tag{27}$$

$$\begin{aligned}\mathbf{h}^{\mathrm{T}} = [&H_{11}(j\omega)\cdots H_{M1}(j\omega)|H_{12}(j\omega)\cdots H_{M2}(j\omega)|\\ &\cdots|H_{1K}(j\omega)\cdots H_{MK}(j\omega)].\end{aligned} \tag{28}$$

The vector **h** thus contains all the elements of the matrix **H** and the vector $\mathbf{r}_l$ consists of the filtered reference signals defined in Eq. (25). The Lth-order vector of Fourier transforms $Z_l(\omega)$ can now be written as

$$\mathbf{z} = \mathbf{R}\mathbf{h}, \tag{29}$$

where **R** is the matrix given by

$$\mathbf{R} = [\mathbf{r}_1^{\mathrm{T}}\mathbf{r}_2^{\mathrm{T}}\cdots\mathbf{r}_L^{\mathrm{T}}]^{\mathrm{T}}. \tag{30}$$

The vector of error signal Fourier transforms $E_l(\omega)$ can now be written as

$$\mathbf{e} = \mathbf{d} + \mathbf{R}\mathbf{h}. \tag{31}$$

A cost function is defined that is given by the sum of the error signal power spectra. Thus

$$J(\omega) = \sum_{l=1}^{L} E[E_l^*(\omega)E_l(\omega)], \tag{32}$$

where the expectation operator $E[\cdot]$ formally implies the ensemble average of the Fourier transforms of signals of infinite duration, although in practice suitable spectral estimates will be made. It follows that

$$J(\omega) = E[\mathbf{e}^{\mathrm{H}}\mathbf{e}], \tag{33}$$

and substitution of Eq. (31) into this expression yields a quadratic function of the type defined by Eq. (5) with the parameters

$$\mathbf{A} = E[\mathbf{R}^{\mathrm{H}}\mathbf{R}], \qquad \mathbf{b} = E[\mathbf{R}^{\mathrm{H}}\mathbf{d}], \qquad c = E[\mathbf{d}^{\mathrm{H}}\mathbf{d}]. \quad (34)$$

Assuming positive definiteness of $\mathbf{A}$, the optimal vector of filter frequency response functions that minimizes the sum of the error sensor power spectra is thus given by $\mathbf{A}^{-1}\mathbf{b}$, and the corresponding minimum value is given by $c - \mathbf{b}^{\mathrm{H}}\mathbf{A}^{-1}\mathbf{b}$.

The case of multiple detected signals and a single error sensor is of particular interest. In this case it follows that the ratio of the minimum value $J_0(\omega)$ of the error sensor power spectrum to its value $J_d(\omega)$ in the absence of control is given by[28]

$$\frac{J_0(\omega)}{J_d(\omega)} = 1 - \eta_{nd}^2(\omega), \quad (35)$$

where $\eta_{nd}^2(\omega)$ is the multiple coherence function[29] between the detection sensor inputs and the error sensor output. In the case of a single detection sensor, this result reduces to the ordinary coherence function. The significance of these coherence functions in the context of active control was pointed out by Ross,[30] and they provide a convenient means for assessing the potential for active control in a practical application. Measurement of these functions can be used to define an upper bound on the performance of an active controller without the necessity for implementing the control system.

7.3 Time-Domain Analysis

The best that can be achieved with a controller that is constrained to be causal can be calculated by conducting the analysis in the discrete-time domain. This has obvious practical application when the controller is implemented as a matrix of finite impulse response (FIR) digital filters, but in undertaking such an analysis the filters are also constrained to have a finite as well as a causal impulse response. Analyses of active control problems can be undertaken in order to determine the structure of optimal causal controllers in continuous time, but this requires the solution of a Wiener Hopf integral equation.[31] Working in discrete time leads naturally to the numerical solution of this equation. The nth sample of the signal due to the secondary sources at the lth error sensor can be written as the convolution

$$\begin{aligned} z_l(n) = {} & \mathbf{h}^{\mathrm{T}}(0)\mathbf{r}_2(n) + \mathbf{h}^{\mathrm{T}}(1)\mathbf{r}_l(n-1) + \cdots \\ & + \mathbf{h}^{\mathrm{T}}(I-1)\mathbf{r}_l(n-I+1), \end{aligned} \quad (36)$$

where it is assumed that each filter in the matrix $\mathbf{H}$ has I coefficients and a composite tap weight vector and a sampled reference signal vector are defined respectively by

$$\begin{aligned} \mathbf{h}^{\mathrm{T}}(i) = {} & [h_{11}(i) \cdots h_{M1}(i) | h_{12}(i) \cdots h_{M2}(i) | \\ & \cdots | h_{1K}(i) \cdots h_{MK}(i)], \end{aligned} \quad (37)$$

$$\begin{aligned} \mathbf{r}_l^{\mathrm{T}}(n) = {} & [r_{l11}(n) \cdots r_{lM1}(n) | r_{l12}(n) \cdots r_{lM2}(n) | \\ & \cdots | r_{l1K}(n) \cdots r_{lMK}(n)]. \end{aligned} \quad (38)$$

A further composite tap weight vector $\mathbf{w}$ can now be defined that consists of all the I tap weights of all the $L \times M$ filters such that

$$\mathbf{w}^{\mathrm{T}} = [\mathbf{h}^{\mathrm{T}}(0)\mathbf{h}^{\mathrm{T}}(1) \cdots \mathbf{h}^{\mathrm{T}}(I-1)]. \quad (39)$$

The Lth-order vector of sampled error signals is now written as

$$\mathbf{e}(n) = \mathbf{d}(n) + \mathbf{R}(n)\mathbf{w}, \quad (40)$$

where the matrix $\mathbf{R}(n)$ is defined by

$$\mathbf{R}(n) = \begin{bmatrix} \mathbf{r}_1^{\mathrm{T}}(n) & \mathbf{r}_1^{\mathrm{T}}(n-1) & \cdots & \mathbf{r}_1^{\mathrm{T}}(n-I+1) \\ \mathbf{r}_2^{\mathrm{T}}(n) & \mathbf{r}_2^{\mathrm{T}}(n-1) & \cdots & \mathbf{r}_2^{\mathrm{T}}(n-I+1) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{r}_L^{\mathrm{T}}(n) & \mathbf{r}_L^{\mathrm{T}}(n-1) & \cdots & \mathbf{r}_L^{\mathrm{T}}(n-I+1) \end{bmatrix}. \quad (41)$$

The relevant cost function is now the sum of the time-averaged error signals given by

$$J = E[\mathbf{e}^{\mathrm{T}}(n)\mathbf{e}(n)]. \quad (42)$$

Substitution of Eq. (40) then shows that the cost function can be written in the quadratic form

$$\begin{aligned} J = {} & \mathbf{w}^{\mathrm{T}}E[\mathbf{R}^{\mathrm{T}}(n)\mathbf{R}(n)]\mathbf{w} + 2\mathbf{w}^{\mathrm{T}}E[\mathbf{R}^{\mathrm{T}}(n)\mathbf{d}(n)] \\ & + E[\mathbf{d}^{\mathrm{T}}(n)\mathbf{d}(n)]. \end{aligned} \quad (43)$$

Again, this function has a unique minimum value associated with an optimal value of the tap weight vector $\mathbf{w}$. In a single-input, single-output controller, the techniques for searching this quadratic performance surface are well established.[32] The application of the method of steepest descent, together with the use of an instantaneous estimate of the gradient of the surface, leads directly to the filtered-x LMS algorithm.[32] With the quadratic function given by Eq. (43), identical arguments can be applied. Thus the gradient of the cost function given by Eq. (43) can be written as

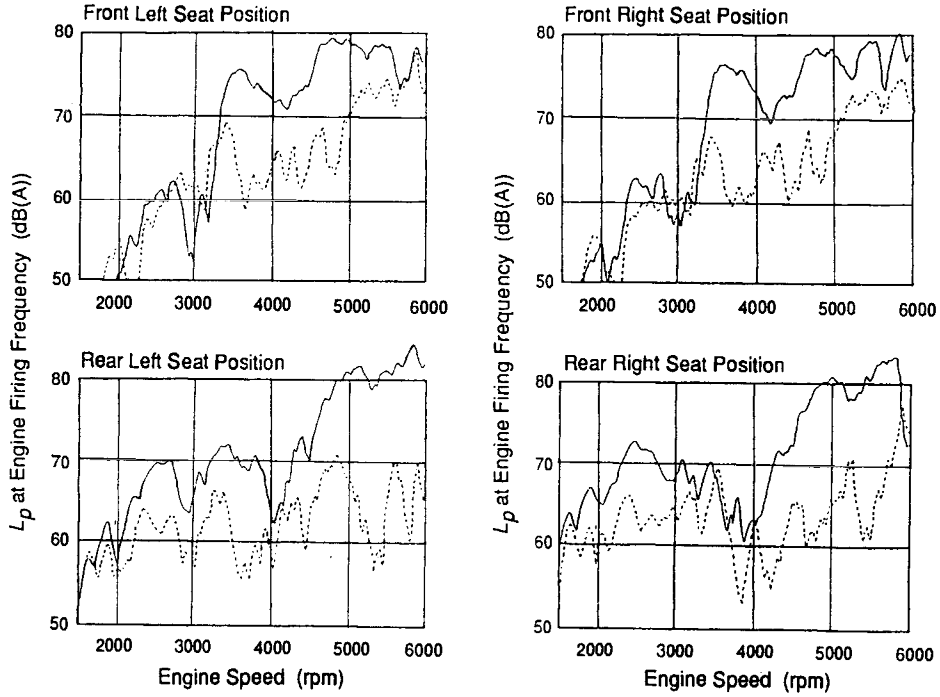


Fig. 11 A-weighted sound pressure level at the engine firing frequency measured at head height in the four seat positions of a 1.1-L small saloon car as it was accelerated hard in second gear with (- - - -) and without (—) a four-loudspeaker, eight-microphone active noise control system in operation. Reprinted with kind permission of Academic Press.

$$\frac{\partial J}{\partial \mathbf{w}} = 2E[\mathbf{R}^{\mathrm{T}}(n)\mathbf{R}(n)\mathbf{w} + \mathbf{R}^{\mathrm{T}}(n)\mathbf{d}(n)], \qquad (44)$$

which, on use of Eq. (40), reduces to

$$\frac{\partial J}{\partial \mathbf{w}} = 2E[\mathbf{R}^{\mathrm{T}}(n)\mathbf{e}(n)]. \qquad (45)$$

The method of steepest descent suggests that the vector $\mathbf{w}$ should be updated by an amount proportional to the negative of the gradient $\partial J/\partial \mathbf{w}$. The use of an instantaneous estimate of this gradient then leads to the tap weight update equation

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \alpha \mathbf{R}^{\mathrm{T}}(n)\mathbf{e}(n), \qquad (46)$$

where α is a convergence coefficient. This has become known as the multiple-error LMS algorithm[33] and has proved most successful when used as the basis for the adaptation of digital filters used in multichannel control systems for the suppression of periodic sound fields in aircraft[34] and automobiles.[35] An illustration of the performance of a typical multichannel active controller used for the suppression of engine-induced noise in automobiles is shown in Fig. 11.

REFERENCES

1. P. A. Nelson and S. J. Elliott, *Active Control of Sound*, Academic, London, 1992.
2. C. R. Fuller, S. J. Elliott, and P. A. Nelson, *Active Control of Vibration*, Academic, London, 1996.
3. M. J. M. Jessel, "Sur les absorbeurs actifs," *Proceedings 6th ICA*, Tokyo, 1968, Paper F-5-6, p. 82.
4. G. D. Malyuzhinets, "An Inverse Problem in the Theory of Diffraction," *Sov. Phys. Doklady*, Vol. 14, 1969, pp. 118–119.
5. A. J. Kempton, "The Ambiguity of Acoustic Sources—A Possibility for Active Control?" *J. Sound Vib.*, Vol. 48, 1976, pp. 475–483.
6. P. E. Doak, "Fundamentals of Aerodynamic Sound Theory and Flow Duct Acoustics," *J. Sound Vib.*, Vol. 28, 1973, pp. 527–561.
7. K. Kido and S. Onoda, "Automatic Control of Acoustic Noise Emitted from Power Transformer by Synthesising Directivity," *Science Reports of the Research Institutes, Tohoku University (RITU), Sendai, Japan*, Series B: Tech-

nology. Part I: Reports of the Institute of Electrical Communication (RIEC) 2397-110, 1972.

8. J. Piraux and B. Nayroles, "A Theoretical Model for Active Noise Attenuation in Three-Dimensional Space," *Proceedings of Inter-Noise '80*, Miami, 1980, pp. 703–706.
9. P. A. Nelson, A. R. D. Curtis, S. J. Elliott, and A. J. Bullmore, "The Minimum Power Output of Free Field Point Sources and the Active Control of Sound," *J. Sound Vib.*, Vol. 116, No. 3, 1987, pp. 397–414.
10. P. A. Nelson, A. R. D. Curtis, S. J. Elliott, and A. J. Bullmore, "The Active Minimization of Harmonic Enclosed Sound Fields, Part I: Theory," *J. Sound Vib.*, Vol. 117, No. 1, 1987, pp. 1–13.
11. P. A. Nelson, A. R. D. Curtis, and S. J. Elliott, "On the Active Absorption of Sound," *Proceedings of Inter-Noise '86*, 1986, pp. 601–606.
12. P. M. Morse and K. U. Ingard, *Theoretical Acoustics*, McGraw-Hill, New York, 1968.
13. R. J. Silcox, "Active Control of Multi-modal Random Sound in Ducts," *Proceedings Institute of Acoustics Spring Meeting*, Southampton, England, 1990.
14. L. J. Eriksson, M. C. Allie, R. H. Hoops, and J. V. Warner, "Higher Order Mode Cancellation in Ducts Using Active Noise Control," *Proceedings of Inter-Noise '89*, Newport Beach, CA, 1989.
15. L. J. Eriksson, M. C. Allie, and R. A. Greiner, "The Selection and Application of an IIR Adaptive Filter for Use in Active Sound Attenuation," *IEEE Trans. Acoust., Speech Signal Process.*, Vol. ASSP-35, No. 4, 1987, pp. 433–437.
16. A. Roure, "Self Adaptive Broad Band Active Sound Control System," *J. Sound Vib.*, Vol. 101, No. 3, 1985, pp. 429–441.
17. C. F. Ross, "An Algorithm for Designing a Broad Band Active Sound Control System," *J. Sound Vib.*, Vol. 80, No. 3, 1982, pp. 373–380.
18. M. A. Swinbanks, "The Active Control of Sound Propagation in Long Ducts," *J. Sound Vib.*, Vol. 27, No. 3, 1973, pp. 411–436.
19. A. D. Pierce, *Acoustics: An Introduction to its Physical Principles and Applications*, McGraw-Hill, New York, 1981.
20. A. J. Bullmore, P. A. Nelson, A. R. D. Curtis, and S. J. Elliott, "The Active Minimization of Harmonic Enclosed Sound Fields, Part II: A Computer Simulation," *J. Sound Vib.*, Vol. 117, No. 1, 1987, pp. 15–33.
21. M. R. Schroeder and K. H. Kuttruff, "On Frequency Response Curves in Rooms. Comparison of Experimental Theoretical and Monte Carlo Results for the Average Frequency Spacing between Maxima," *J. Acoust. Soc. Am.*, Vol. 34, No. 1, 1962, pp. 76–80.
22. S. J. Elliott, P. Joseph, A. J. Bullmore, and P. A. Nelson, "Active Cancellation of a Point in a Pure Tone Diffuse Sound Field," *J. Sound Vib.*, Vol. 120, No. 1, 1988, pp. 183–189.
23. H. F. Olson and E. G. May, "Electronic Sound Absorber," *J. Acoust. Soc. Am.*, Vol. 25, 1953, pp. 1130–1136.
24. P. D. Wheeler, "Voice Communications in the Cockpit Noise Environment—The Role of Active Noise Reduction," Ph.D. Thesis, University of Southampton, England, 1986.
25. C. Carme, "Absorption Acoustique Active Dans Les Cavites," Ph.D. Thesis, l'Universite d'Aix-Marseille II, 1987.
26. W. K. W. Hong, Kh. Eghtesadi, and H. G. Leventhall, "The Tight-coupled Monopole (TCM) and Tight-coupled Tandem (TCT) Attenuators: Theoretical Aspects and Experimental Attenuation in an Air Duct," *J. Acoust. Soc. Am.*, Vol. 81, 1987, pp. 376–388.
27. M. C. J. Trinder and P. A. Nelson, "Active Noise Control in Finite Length Ducts," *J. Sound Vib.*, Vol. 89, No. 1, 1983, pp. 95–105.
28. P. A. Nelson, T. J. Sutton, and S. J. Elliott, "Performance Limits for the Active Control of Random Sound Fields from Multiple Primary Sources," *Proceedings of IOA Spring Conference*, Southampton, England, 1990.
29. D. E. Newland, *An Introduction to Random Vibrations and Spectral Analysis*, 2nd ed., Longman Scientific and Technical, New York, 1984.
30. C. F. Ross, "Active Control of Sound," Ph.D. Thesis, University of Cambridge, England, 1980.
31. P. A. Nelson, J. K. Hammond, P. Joseph, and S. J. Elliott, "Active Control of Stationary Random Sound Fields," *J. Acoust. Soc. Am.*, Vol. 87, No. 3, 1990, pp. 963–975.
32. B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1985.
33. S. J. Elliott, I. M. Stothers, and P. A. Nelson, "A Multiple Error LMS Algorithm and Its Application to the Active Control of Sound and Vibration," *IEEE Trans. Acoust., Speech Signal Process.*, ASSP-35, No. 10, 1987, pp. 1423–1434.
34. S. J. Elliott, P. A. Nelson, I. M. Stothers, and C. C. Boucher, "In-flight Experiments on the Active Control of Propeller-induced Cabin Noise," *J. Sound Vib.*, Vol. 140, 1990, pp. 219–238.
35. S. J. Elliott, I. M. Stothers, P. A. Nelson, A. M. McDonald, D. C. Quinn, and T. Saunders, "The Active Control of Engine Noise inside Cars," *Proceedings of Internoise 88*, Avignon, France, 1988, pp. 987–990.

85

NOISE SOURCES AND PROPAGATION IN DUCTED AIR DISTRIBUTION SYSTEMS

HOWARD KINGSBURY

1 INTRODUCTION

This chapter deals with the generation, propagation, and control of noise in a ducted air distribution system. First, fan noise control is considered, which is considered with reference to a simple system, with complications added later. The discussion refers to octave band noise generation and control. Included is discussion of both "natural attenuation," that is, the noise reduction that occurs due to the transmission of sound through the duct walls, the division of noise due to duct branches, and the effect of duct fittings. Also treated are the effects of porous duct linings and packaged attenuation units (silencers). Consideration is also given to the conversion of sound power levels to sound pressure levels in the room served by the fan system. The methods and effects of varying the air volume delivered by the fan is discussed. Second, turbulent flow noise generation and control in the ducts are covered. The generation and control of "break-out noise," that is, noise that is radiated through the duct wall, is discussed, as are the differences between round and rectangular duct in the control of this noise. The effects of volume control dampers located near outlets are included. Special consideration is given to rooftop units, where turbulent flow is a common cause of noise generation. Third, noise generation and control in the outlet devices are considered. Both constant-air-volume and variable-air-volume (VAV) types of terminals are included, including so-called fan-powered terminals. Information on the sound generated by fans is in Chapter 98.

2 FAN NOISE CONTROL IN DUCTED SYSTEMS

Understanding of noise control in a ducted air distribution system can best be developed by beginning with a simple system and then adding the complications. Figure 1 shows a ducted system reduced to its barest essentials. It is a constant-volume system, with a single fan running at constant speed, enclosed in a plenum, both supplying and returning the air from the room being served. Eliminated from this system are such accessories as heating or cooling coils, filters, and so on. Usually, the most critical part of the system is the room closest to the fan serving that system unless other rooms downstream require lower sound levels.

The recommended noise control procedure is as follows, working in octave bands:

1. Determine the sound power output of the fan(s) that is emitted into the duct system.
2. Determine the so-called natural attenuation of the duct system. While doing so, determine if the sound generated by fittings and take-offs may add to that being transmitted down the duct.
3. Convert from the sound power level at the outlet to the resulting sound pressure level in the space being served (so-called room factor or space effect) and compare it with the chosen sound level criterion for that space.
4. Add any required sound attenuation measures to the system needed to meet the criterion in 3.
5. Determine if sound transmission through the duct wall (break-out noise) may be a problem and, if so, apply proper sound control procedures.

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker
Copyright © 1997 John Wiley & Sons, Inc.

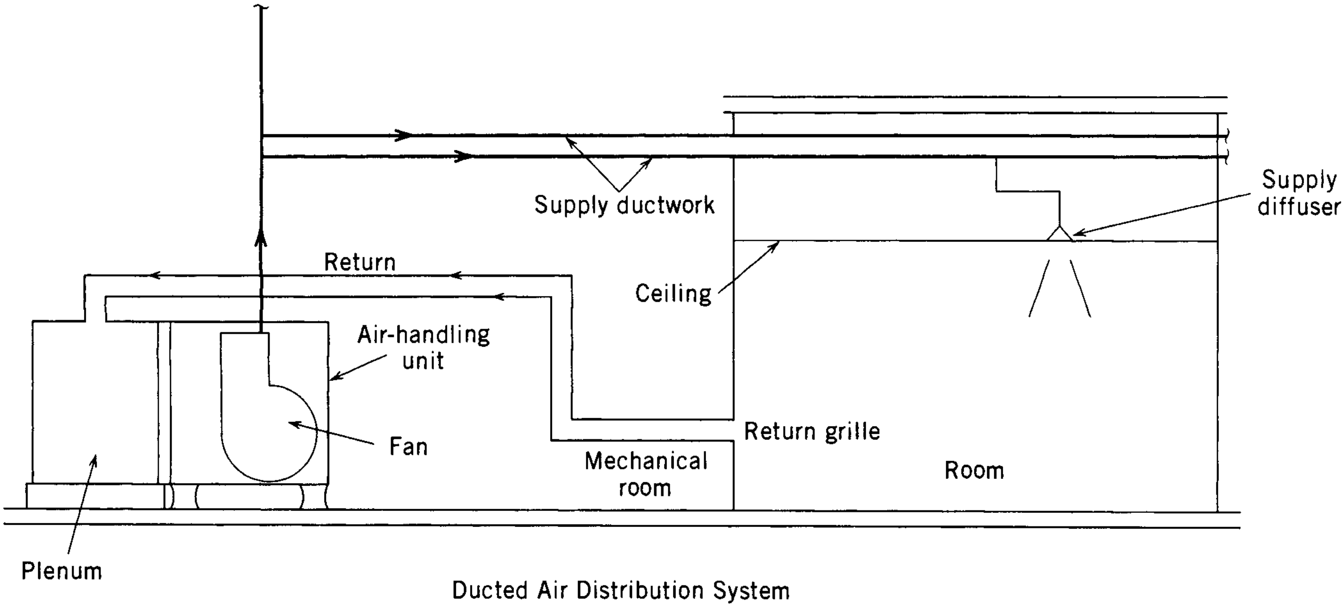


Fig. 1 Ducted air distribution system.

6. Repeat the entire procedure for the return side of the system.

In addition to these considerations of fan noise, note that the final sound level in the room will be a combination of the noise from the fan and noise from the chosen outlet device(s), which will be discussed later.

2.1 Natural Attenuation

The first aspect of noise control in a duct distribution system is the so-called natural attenuation, which has four components. First, some sound is radiated through the duct wall, causing the duct wall to vibrate; this is sensed as a reduction of energy inside the duct (see Table 1).[1] (Round and flat oval ducts show less of this effect than rectangular ducts due to their shape. For round ducts, use an attenuation of 0.1 dB/m up to 1000 Hz, rising to 0.4 dB/m at 4000 Hz.) A second mechanism is the division of energy between the main duct and the take-offs. In the absence of specific information to the contrary, it is assumed the energy division is proportional to the ratio of the branch duct area to that of the main duct (see Table 2). A third common loss occurs at the end of the duct, where it enters the room. This is called end-reflection loss (some of the sound energy is reflected back into the duct system) and is most pronounced when there is a sudden transition from the duct to the room. Diffusers reduce this effect; it is suggested that it be reduced by 6 and 4 dB in the 63- and 125-Hz octave bands for duct sizes 300 mm and under when diffusers are on the end of the duct (see Table 3).[1] Finally, a fourth component is the conversion from sound power in the duct system to the sound pressure in the room (the room factor noted above). For most practical rooms, the classical equation (Chapter 91), which assumes both a direct and reverberant field, does not apply. A room having an acoustical

TABLE 1 Attenuation of Unlined Sheet Metal Ducts, (db/unit length) as Function of P/A by Octave Band Center Frequency

P/A (mm/mm^2)	Attenuation (dB/m) 63 Hz	125 Hz	≥250 Hz
Over 0.012	0.05	1.0	0.33
0.012–0.005	1.0	0.33	0.33
Under 0.005	0.33	0.33	0.33

Note: If duct is externally insulated, double above listed attenuations; P = duct perimeter; A = cross-sectional area.

Source: Ref. 1, p. 42.11.

TABLE 2 Division of Sound Power at Branch Take-offs

Ratio	dB Reduction
1.0	0
0.5	3
0.25	6
0.10	10

Note: Ratio = branch duct/continuing duct area.

TABLE 3 Duct End Reflection Loss by Octave Band Center Frequency

Mean Duct Width (mm)	Reflection Loss (dB)				
	63 Hz	125 Hz	250 Hz	500 Hz	1000 Hz
150	18	12	8	4	1
250	14	9	5	1	0
500	9	5	1	0	0
1000	4	1	0	0	0

Note: If duct terminates in a diffuser, deduct 6 dB (63 Hz) and 4 dB (125 Hz) for duct sizes under 300 mm, 3 dB (63 Hz) for duct sizes over 300 mm. Do *not* apply to linear diffusers or diffusers tapped directly into primary duct work.

Source: Ref. 1, p. 42.16.

ceiling and in which the ceiling height is significantly less than either of the horizontal dimensions, the Schultz equation (Chapter 91) should be used for this conversion.

2.2 Lined Duct Attenuation

In many systems additional attenuation is required. This may be obtained by lining the interior surfaces of the duct with sound-absorbing material or by use of packaged attenuators or both. Duct lining serves two purposes: It is a thermal insulant and a sound absorber. While 13 mm thickness may be sufficient for thermal reasons, for acoustical purposes the thickness should be a minimum of 25 mm, and 50 mm is often desirable. This is due to the minimal low-frequency attenuation provided by thin linings, and it is in the low-frequency region that most duct system noise control problems occur.

The acoustical characteristics of glass fiber duct lining in rectangular ducts have been rather extensively investigated; the same is not true for round ducts and in the absence of further experimental evidence, Table 4[1] should be used for round-duct attenuation.

The attenuation of fiber glass lined rectangular duct sections may be calculated using Eqs. (1) and (2).[1]

For frequencies below 800 Hz

$$\text{Total attenuation} = \frac{t^{1.08} h^{0.356} (P/A)/L f^{(1.17 + Kd)}}{K_1 d^{2.3}}, \quad (1)$$

where $K_1 = 5.46 \times 10^{-3}$, $K = 0.0119$, h is the smaller inside dimension of the duct in millimetres, f is frequency in hertz, P is the inside duct perimeter in millimetres, A is the inside area of the duct in square millimetres, and L is the duct length in metres.

Note that since the attenuation is strongly frequency dependent, it is suggested that low-frequency octave band attenuation be evaluated for the one-third-octave band below the octave band center frequency of interest, that is, 100 Hz for the 125-Hz octave band. Also, add the attenuation of unlined duct (Table 1) to the attenuation provided by this calculation for the 63- and 125-Hz octave bands.

For frequencies above 800 Hz

$$\text{Total attenuation} = K_4 \frac{(P/A) L f^{(K_5 - 1.611 \text{ of } P/A)}}{w^{2.5} h^{2.7}}, \quad (2)$$

where w is the larger inside dimension of duct in millimetres, $K_4 = 3.32 \times 10^{18}$, and $k_5 = 3.79$.

Note that the above equation may be used for straight duct lengths of 3 m or less; for straight lengths (i.e., no elbows) longer than 3 m, limit the total attenuation to 40 dB. The limit of 40 dB for any straight duct length applies to both equations, since structural flanking through the duct wall imposes the upper limit. Further, do not use either equation when the duct is constructed of glass fiber with a lightweight, low-transmission-loss cov-

TABLE 4 Insertion Loss for Lined Circular Ducts with 25- or 50-mm-Thick Lining by Octave Band Center Frequency

	Insertion Loss, dB/m											
	63 Hz		125 Hz		250 Hz		500 Hz		1000 Hz		2000 Hz	
Diameter (mm)	25 mm	50 mm	25 mm	50 mm	25 mm	50 mm	25 mm	50 mm	25 mm	50 mm	25 m	50 mm
150	1.3	1.8	1.9	2.6	3.1	4.5	5.0	7.4	7.2	7.2	7.6	7.6
300	0.76	1.4	1.5	2.2	2.7	4.1	4.8	7.2	7.2	7.2	6.3	6.3
600	0.23	0.83	0.83	1.5	1.9	3.3	4.2	6.6	5.6	5.6	4.1	4.1
1200	0	0.36	0	0.50	0.59	2.1	2.1	4.5	.86	.86	1.1	1.1

Source: Ref. 1, p. 42.14.

TABLE 5 Insertion Loss for Lined, Square Elbows with Turning Vanes by Octave Band Center Frequency

	Insertion Loss (dB)					
Duct Width (mm)	63 Hz	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz
125–250	0	0	1	2	3	4
260–510	0	1	2	3	4	6
520–1020	1	2	3	4	5	6
1030–2030	2	3	3	5	6	8

Note: Duct width is dimension in plane of the turn.
Source: Ref. 1, p. 42.15.

ering, such as a metal foil. Sound transmission through the covering limits the total attenuation, since the sound reenters the duct downstream.

2.3 Lined Rectangular Elbows

Elbows with turning vanes also provide worthwhile attenuation. Round elbows, lined or unlined, provide relatively little attenuation, in general 5 dB or less. Table 5 provides typical values for rectangular elbows.[1]

2.4 Passive Attenuators or Silencers

These items are available in a variety of sizes and lengths. Commonly available lengths of rectangular silencers are 0.9, 1.5, 2.1, and 3.1 m. The length of most circular cross-sectional silencers is a function of their diameter. In the selection of silencers, three interrelated factors must be considered: (1) insertion loss, (2) pressure drop, and (3) self-generated noise. The primary controlling element is the airflow velocity through the face area of the silencer. Increasing velocity increases both the pressure drop and the self-noise; it reduces the insertion loss slightly. Manufacturers' literature should be consulted for details on these items. A recognized test method for determining these properties is ASTM E-477, which measures the insertion loss by the duct-to-reverberant-room method with and without airflow.

2.5 Active Sound Attenuation

Recent developments in computer technology have led to the development of phase-canceling systems having application to low-frequency noise in ducted systems. Both random noise and tones are accounted for, with an upper frequency limit of about 400 Hz and insertion losses typically of 20–30 dB.

2.6 Attenuation of Noise in Lined Plena

The noise reduction offered by lined fan plena, as shown in Fig. 2, should be taken into account. The following approximate expression may be used to calculate the effects of such a lined plenum[2]:

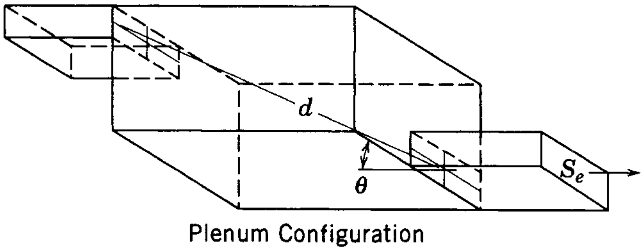


Fig. 2 Diagram of sound-absorbing plenum.

$$\text{Attenuation} = 10\log\left[S_e\left(\frac{\cos\Theta}{2\Pi d^2} + \frac{1-\alpha}{\alpha S_w}\right)\right]^{-1}, \quad (3)$$

where α is the absorption coefficient of lining at octave band center frequencies, S_e the plenum exit area in square metres, d the slant distance between entrance and exit in metres, S_w the plenum wall area in square metres, and Θ the angle of incidence at exit, that is, the angle that d makes with the normal to the exit. The results of this calculation are accurate to within a few decibels for frequencies where the wavelength is much less than the plenum dimensions. At low frequencies (63 and 125 Hz) it is conservative by 5–10 dB due to sound reflections at the entrance and exit of the plenum.

3 AIR TERMINAL DEVICES

In most ducted systems, a device is placed at the end of the duct to deliver the air to the room. These are most commonly diffusers, grilles, or registers and may contain volume control devices. Since they radiate sound directly into the room, they must be selected to meet the room noise level criterion. For the most part, their noise emissions occur in the octave bands above 250 Hz. Manufacturers' literature should be consulted for appropriate

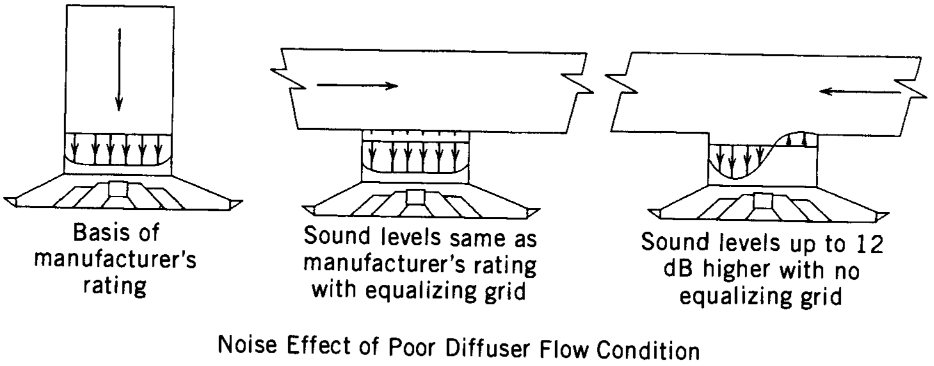

Basis of manufacturer's rating
Sound levels same as manufacturer's rating with equalizing grid
Sound levels up to 12 dB higher with no equalizing grid
Noise Effect of Poor Diffuser Flow Condition

Fig. 3 Noise effect of poor diffuser flow condition. (From Ref. 1, p. 42.19.)

sound ratings. Care should be taken that the conditions for which the data are presented match the application at hand and any adjustments made as necessary. The noise radiated by a terminal device is dependent not only on the volume and velocity of the air flow through the element but also on the upstream flow conditions. Manufacturer's data are taken under essentially ideal flow conditions; poor flow conditions may add 10–15 dB to the catalog data. A variety of upstream airflow conditions and associated additional sound levels are shown in Fig. 3.[1] Frequently, flexible duct is used to connect the terminal device to the branch duct. The length of this flexible duct should be kept as short as possible to avoid additional flow noise, with 1 m being a suggested maximum. Misalignment of the flexible duct can cause increased sound levels, as shown in Fig. 4.[1]

Volume control dampers may be another source of increased noise levels, typically if located close to the outlet. The noise they create is a function of the additional pressure drop they introduce into the system; the additional noise is 20 log of the ratio of the additional pressure drop to that of the diffuser itself. In acoustically critical spaces, volume control dampers should be located well back in the duct system (5–10 duct diameters from the outlet), with the lined duct between the damper and the outlet.

3.1 Variable-Air-Volume Systems

For energy conservation reasons, many current air conditioning systems modulate the airflow to the conditioned space to match the required heating or cooling load, as it varies throughout the day and year. These VAV systems have two distinguishing characteristics from the acoustical point of view: (1) a terminal device to modulate the airflow and (2) some method of varying the flow volume from the fan.

3.2 Variable-Air-Volume Terminals

In its simplest form, the terminal device consists of an automatically operating damper in a sheet metal box. It is commonly connected directly to the supply duct and with one or more outlets feeding diffusers. Noise is created by the damper as a function of the pressure drop

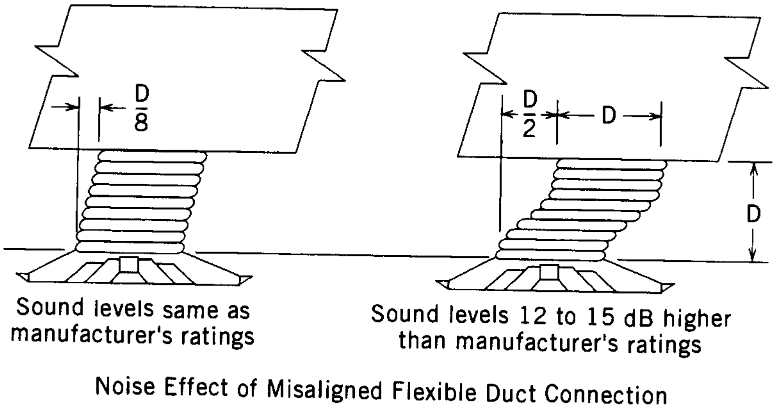

D/8
D/2
D
D
Sound levels same as manufacturer's ratings
Sound levels 12 to 15 dB higher than manufacturer's ratings
Noise Effect of Misaligned Flexible Duct Connection

Fig. 4 Effect of proper and improper alignment of flexible duct connector. (From Ref. 1, p. 42.23.)

across the damper. This noise can be radiated by two paths: (1) directly through the wall of the unit and (2) downstream through the connected duct(s). This noise is frequently controlled by a section of lined duct between the box and the outlet(s). A more difficult control case is presented when the outlet and VAV terminal are combined. In some situations, the casing radiated sound must be controlled. In most cases, such terminal units are located above acoustical panel ceilings, which have limited sound transmission loss properties and may contain acoustically transparent areas such as return air grilles and light fixtures. In general, VAV boxes that handle more than about 0.8 m^3/s should not be located over acoustical environments such as offices; they should be located over more noncritical spaces such as corridors, storage areas, or toilet rooms. In some cases, where such acoustically less sensitive spaces are not available and large box sizes must be used, it may be necessary to increase the transmission loss characteristics of the ceiling by incorporating a layer of 12.5-mm gypsum board between the unit and the acoustical ceiling or encasing the box with a 25-mm layer of glass fiber and 12.5-mm gypsum board.

3.3 Fan-Powered VAV Terminals

A special case is a VAV terminal that incorporates a fan to mix the air from the plenum above the ceiling with the primary air. These are called fan-powered terminals (FPTs) and are of two kinds: one in which the fan cycles on and off, as required, and a second in which the fan runs full time. Since the fan in the first goes on and off, the occupants are more aware of the control unit sound. The noise emitted by these units consists of both the noise radiated by the casing and the fan and that through the room air outlet. By reason of the fan contribution to the total direct radiated noise, units handling more than about 0.8 m^3/s commonly produce noise levels in the space below corresponding to a noise criteria (NC) rating of 40 or above and hence should not be located over spaces requiring a lower sound level. Some possible noise control measures for these units include, in addition to moving them to more remote locations, keeping them at least 2 m from any open return grilles in the ceiling, installing sound attenuators if offered as an option by the manufacturer, or field fabricating lined elbows for installation on the induced air opening of the FPT. Consult manufacturers' literature for appropriate information on the noise levels produced by the above units and how to determine the noise levels in the occupied space served by them. Data on these units are taken in accordance with the Air-Conditioning & Refrigerating Institute (ARI) standard 880-89 (estimation of the sound level in the occupied space by ARI 885).

3.4 Fan Volume Control

As the airflow volume through the terminals varies to meet the required heating or cooling load, the volume of air produced by the fan must also vary. There are three basic methods of doing so: (1) use of inlet vanes on the fan, (2) motor speed controls, and (3) varying the pitch of the fan blades:

Inlet vanes are positioned at the inlet to the fan scroll and close down to reduce airflow. They add resistance to the system and, depending on the fan type, may cause significant additional low-frequency fan sound, which must be dealt with further downstream. Since resistance to flow is added, there is also energy usage inefficiency when the vanes are in their partially open position.

Variable-speed controls reduce the speed of the fan and thus the volume flow; they also reduce the fan noise generation. Variable-speed controls have a significant energy savings over inlet vanes; this should be taken into account in the selection process.

Variable-pitch blades are available for some axial flow fans; this also results in quieter operation at reduced flow.

4 SOUND TRANSMISSION THROUGH DUCT WALLS

Sound radiated through duct walls may be a problem if not sufficiently attenuated before the duct passes over or through occupied space. This is most frequently a problem close to the fan room, where the noise in the duct may be composed of both that from the fan and that aerodynamically generated due to poor flow conditions. Typically, this noise is concentrated in the octave bands below 250 Hz and may extend down through the 31.5-Hz band. It is commonly called *breakout noise*, and its level is dependent on the noise level within the duct, the duct cross section, and the length of the duct in or over the occupied space. The transmission loss characteristics of rectangular, flat oval, and round ducts have been determined; representative values for round and rectangular ducts are shown in Table 6.[1] The transmission loss of rectangular ducts does not vary significantly as a function of duct size or construction; the transmission loss of round ducts does vary as a function of the duct diameter and values for several typical diameters are included. Note that at low frequencies the transmission loss of round ducts is markedly higher than for rectangular ones. Using round ducts is an excellent method of controlling low-frequency breakout noise. Since the noise is

TABLE 6 Transmission Loss_{out} for Rectangular and Round Duct by Octave Band Center Frequency (Hz)

	63Hz	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz
Rectangular Sizes (300–1200 mm)	20	23	26	29	32	38
Round Duct Dia. (mm)						
200	> 48	> 65	> 75	> 72	56	56
350	> 48	> 53	55	33	34	35
650	> 48	> 53	36	32	32	28
800	> 48	42	38	25	26	24

Source: Ref. 1, p. 42.23.

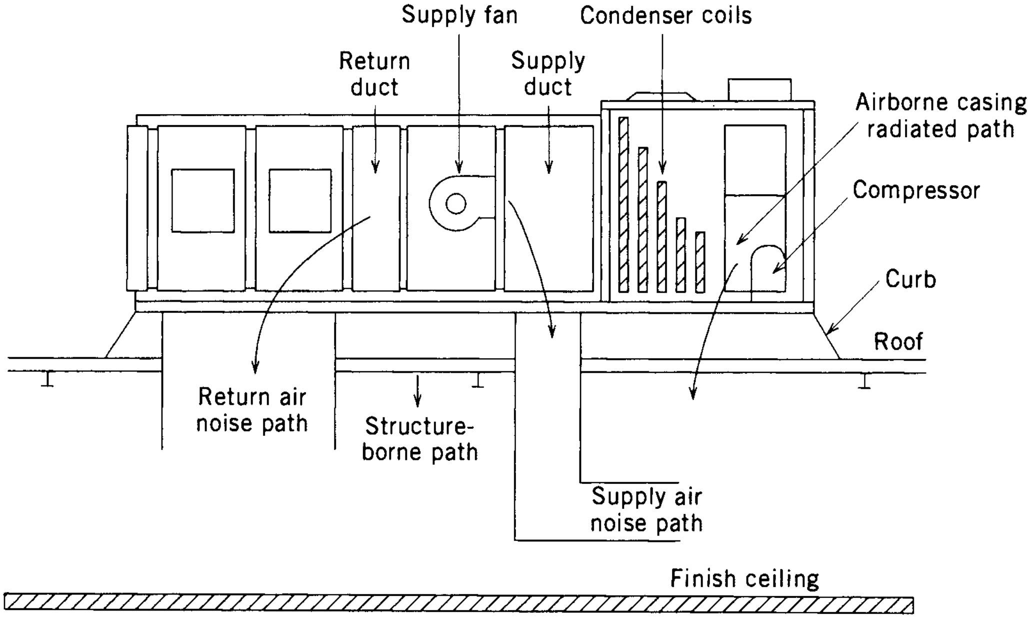
better contained, additional duct system attenuation may be needed downstream for appropriate control of room sound levels. The transmission loss of flat oval duct is intermediate between rectangular and round duct, but due to the relatively large flat surfaces, it tends to follow the same characteristics as that of rectangular duct, that is, lower values at low frequencies.

5 TURBULENT AIRFLOW NOISE CONTROL

Although fans are usually the major source of noise in a duct system, aerodynamic noise may be generated at elbows, turns, take-offs, silencers, and other duct elements. The generated noise levels depend on the element geometry as well as the airflow velocity. Since the intensity of flow generated sound is proportional to between the fifth and sixth power of the local velocity in the duct, the best methods of control are to (1) keep the duct velocities as low as possible and (2) avoid closely spaced elements (turns, take-offs) or abrupt changes in duct area that can contribute to turbulent flow. In general, if the flow velocities are below about 10 m/s and elements can be separated by about three to five duct diameters, most aerodynamically induced noise control problems are minimized.

6 ROOFTOP UNITS

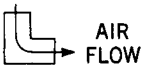
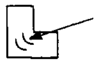
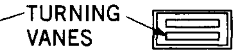
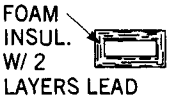
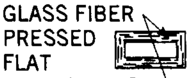
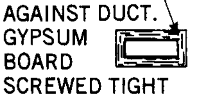
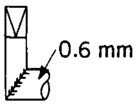
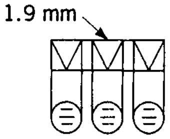
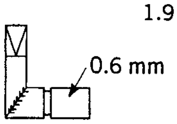
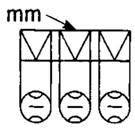
Rooftop units (packaged, factory assembled units containing fans and associated heating and cooling equip-



Packaged Rooftop Air Conditioning Unit

Fig. 5 Noise paths in typical rooftop installations. (From Ref. 1, p. 42.27.)

TABLE 7 Duct Breakout Insertion Loss: Potential Low-Frequency Improvement over Bare Duct Elbow

Discharge Duct Configuration of Horizontal Supply Duct	Duct Breakout Transmission Loss at Low Frequencies (dB)				
	63 Hz	125 Hz	250 Hz	Side View	End View
Rectangular duct: no turning vanes (reference).	0.0	0.0	0.0		
Rectangular duct: two dimensional turning vanes.	0.0	0.5	1.0		
Rectangular duct: wrapped with foam and lead.	4.0	3.0	4.5	SEE END VIEW	
Rectangular duct: wrapped with glass fiber, one layer of gypsum board.	4.0	7.0	5.5	SEE END VIEW	
Rectangular duct: wrapped with glass fiber, two layers of gypsum board.	7.5	8.5	9.0	SEE END VIEW	
Rectangular to multiple drop: round mitered with turning vanes, three parallel round supply ducts.	17.5	11.5	13.0		
Rectangular to multiple drop: round mitered elbowd with turning vanes, three parallel round lined double wall supply ducts	18.0	13.0	16.0		

Source: Ref. 1, p. 42.28.

ment) post important noise control problems. They are frequently mounted on curbs supported by lightweight roof structure directly above occupied space. Because of their proximity to the building interior, they should be located over acoustically noncritical spaces. The primary noise paths that need consideration, as shown in Fig. 5, are[1] (1) the casing-radiated noise, (2) the supply and return air ducts, and (3) structure-borne noise.

6.1 Casing-Radiated Noise

This noise is best controlled by making certain the roof structure inside the curb has adequate mass to block the sound being radiated by the fan casing and compressor, if included. This construction may consist of concrete or multiple layers of gypsum board. Sealing around the perimeter of the curb and around the supply and return duct penetrations is important. The supply and return ducts are commonly directly above the ceiling of the occupied space, and often there is insufficient space for easy application of adequate noise control measures. On the supply side, low-frequency breakout noise is a common problem. Control methods consist of proper configuration of the fan outlet and ducts. A common problem is the duct elbow between the fan outlet and the horizontal duct above the ceiling. Round elbows and round duct are a very important way of controlling this noise. See Table 7.[11]

6.2 Structure-Borne Noise

This noise is controlled by proper vibration isolation. Internal isolation of fans in units without compressor/condenser units will usually be sufficient if the isolators have adequate deflection; steel springs of 50–75 mm deflection are usually required. Additionally, several manufacturers can supply curbs with high deflection springs for such isolation.

An alternate noise control technique is to raise the unit above the roof on a structure (commonly called dunnage). In this configuration, the ducts can be run above the roof before entering the building, creating the opportunity to include noise control measures before the ducts enter the building; casing-radiated noise through the bottom of the unit is also less difficult. Vibration transfer to a lightweight roof is avoided, since the dunnage is usually supported from heavier roof-supporting structure.

REFERENCES

1. *ASHRAE Handbook, HVAC Applications*, ASHRAE, Atlanta, GA, 1991.
2. R. J. Wells, "Acoustical Plenum Chambers," *Noise Control*, July, 1958.

86

SOUND POWER LEVEL PREDICTIONS FOR INDUSTRIAL MACHINERY

ROBERT D. BRUCE AND CHARLES T. MORITZ

1 INTRODUCTION

Procedures for calculating the sound power level or sound pressure level of industrial machinery, such as boilers, fans, motors, pumps, and turbines, are presented in this chapter. These procedures can be used for modeling the sound levels in a space or for developing purchase specifications for new equipment. With any project, acoustical data for individual equipment, specifically sound pressure or sound intensity measurements and sound power level calculations in accordance with recognized standards, should be obtained. Many manufacturers provide the estimated sound power level or the measured sound pressure level at 1 m from their equipment as well as offer special low-noise options. If information from the manufacturer is unavailable, efforts should be made to measure a similar unit in operation at an installation. If this is not possible or if an estimate for a preliminary study is required, then the material in this chapter can be used. The equations presented are either entirely based on measured data or are semiempirical and tend to be conservative, usually predicting higher sound levels than are measured in the field. For some equipment, a large database has not yet been established so estimates of the sound power level for similar equipment are presented.

2 POWER SOURCES

2.1 Boilers[1]

Main Steam Boilers The sound power level for main steam boilers (between 125 and 800 MWe) can be calculated using the equation

$$L_w = 84 + 15 \log \text{MWe}, \qquad (1)$$

where L_w is the overall sound power level and MWe is the electrical generating rating of the unit.

The octave band sound power levels can be obtained by subtracting the values shown in Table 1.

Auxiliary Boilers The noise produced by auxiliary boilers is often due primarily to the blower and the burner and not the walls of the boiler. An estimate of the sound power level for auxiliary boilers between 50 and 2000 boiler horsepower (bhp) can be calculated using the equation

$$L_w = 95 + 4 \log \text{bhp}, \qquad (2)$$

where 1 bhp = 15 kg steam/h. The octave band sound power levels can be obtained by subtracting the values shown in Table 1.

2.2 Electric Motors

Motors under 750 kW[2] For totally enclosed fan-cooled (TEFC) motors, the sound power level can be cal-

---

*The information from Ref. 1 is reprinted by permission of the Edison Electric Institute (EEI) from EEI's *Electric Power Plant Environmental Noise Guide*, 2nd ed., Copyright 1984, all rights reserved. Neither EEI nor any member of EEI (a) makes any warranty, express or implied, with respect to the use of this information or that such use may not infringe privately owned rights or (b) assumes any liability with respect to the use of or for damages resulting from the use of this information.

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker
Copyright © 1997 John Wiley & Sons, Inc.

TABLE 1 Octave Band and A-Weighted Sound Power Level Adjustments

Source	31.5	63	125	250	500	1000	2000	4000	8000	A
Main steam boiler	4	5	10	16	17	19	21	21	21	12
Auxiliary boiler	6	6	7	9	12	15	18	21	24	9
TEFC motors under 750 kW	14	14	11	9	6	6	7	12	20	1
Drip-proof motors under 750 kW	9	9	7	7	6	9	12	18	27	4
Gas turbine casing	10	7	5	4	4	4	4	4	4	2
Gas turbine exhaust	12	8	6	6	7	9	11	15	21	4
Gas turbine intake	19	18	17	17	14	8	3	3	6	0
Reciprocating engines (<600 rpm)	12	12	6	5	7	9	12	18	28	4
Reciprocating engines (600–1500 rpm)	14	9	7	8	7	7	9	13	19	3
Reciprocating einges (600–1500 rpm with blower)	22	16	18	14	3	4	10	15	26	1
Reciprocating engines (>1500 rpm)	22	14	7	7	8	6	7	13	20	2
Reciprocating engine air inlet (turbocharged)	4	11	13	13	12	9	8	9	17	3
Reciprocating engine exhaust	5	9	3	7	15	19	25	35	43	12
Steam turbines	11	7	6	9	10	10	12	13	17	5
Steam turbine generator units	9	3	5	10	14	18	21	29	35	12
Transformers	−3	3	5	0	0	−6	−11	−16	−23	0
Centrifugal compressor casing	10	10	11	13	13	11	7	8	12	2
Centrifugal compressor air inlet	18	16	14	10	8	6	5	10	16	0
Rotary and reciprocating air compressors	11	15	10	11	13	10	5	8	15	2
Feed pumps (1000–9000 kW)	11	5	7	8	9	10	11	12	16	4
Feed pumps (9500–18,000 kW)	19	13	15	11	5	5	7	19	23	1
Centrifugal fan	11	9	7	8	9	9	13	17	24	[a]
Centrifugal fan casing	3	6	7	11	16	18	22	26	33	[a]
Gas recirculation fan casing	10	7	4	7	18	20	25	27	31	12
Axial-flow fan	11	10	9	8	8	8	10	14	15	3
Generators	11	8	7	7	7	9	11	14	19	4
Motor-driven pumps	13	12	11	9	9	6	9	13	19	2
Mechanical-draft cooling towers (full speed)	9	6	6	9	12	16	19	22	30	10
Mechanical-draft cooling towers (half speed)	9	6	6	10	10	11	11	14	20	5
Chiller with reciprocating compressor	—	19	11	7	1	4	9	14	—	0
Chiller with rotary-screw compressor	25	19	15	3	6	10	15	20	22	5
Centrifugal chillers, internal geared	—	8	5	6	7	8	5	8	—	0
Centrifugal chillers, direct drive	—	8	6	7	3	4	7	12	—	0
Centrifugal chillers, >1000 tons	—	11	11	8	8	4	6	13	—	0
Diesel powered, mobile equipment	—	11	6	3	8	10	13	19	25	5

[a] Dependent on blade passage frequency.

culated using the following equations:

$$L_w = 17 + 17 \log \text{kW} + 15 \log \text{rpm} + 10 \log A \quad \text{(under 40 kW)}, \tag{3}$$

$$L_w = 28 + 10 \log \text{kW} + 15 \log \text{rpm} + 10 \log A \quad \text{(over 40 kW)}, \tag{4}$$

where kW is the nameplate motor rating, rpm is the speed at which the motor is operating, and *A* is the conformal surface area (in square metres) at 1 m from the motor. For TEFC motors between 300 and 750 kW, use the value 300 kW in Eq. (4).

The octave band sound power levels can be obtained by subtracting the values shown in Table 1.

For drip-proof motors, the sound power level can be calculated using the following equations:

$$L_w = 12 + 17 \log \text{kW} + 15 \log \text{rpm} + 10 \log A \quad \text{(under 40 kW)}, \tag{5}$$

$$L_w = 23 + 10 \log \text{kW} + 15 \log \text{rpm} + 10 \log A \quad \text{(over 40 kW)}. \tag{6}$$

For drip-proof motors between 300 and 750 kW, use the value 300 kW in Eq. (6).

The octave band sound power levels can be obtained by subtracting the values shown in Table 1.

TABLE 2 Octave Band and A-Weighted Sound Power Levels

Source	31.5	63	125	250	500	1000	2000	4000	8000	*A*
Electric motor 750–4000 kW (1800–3600 rpm)	94	96	98	98	98	98	98	95	88	104
Electric motor 750–4000 kW (1200 rpm)	88	90	92	93	93	93	98	88	81	101
Electric motor 750–4000 kW (900 rpm)	88	90	92	93	93	96	96	88	81	101
Electric motor 750–4000 kW (<720 rpm)	88	90	92	93	93	98	92	83	75	100
Electric motor 750–4000 kW (250 and 400 rpm vertical)	86	87	88	88	88	98	88	78	68	99
1200 kW Triplex mud pump	102	104	108	105	104	99	93	86	83	105
Mud reconditioning degasser	78	83	82	87	94	94	95	88	81	99
Drawworks	96	104	109	112	112	108	105	102	96	113
Top drive	98	97	100	110	105	100	100	96	82	108
Natural draft cooling tower, rim noise	—	—	105	104	106	108	110	112	110	117
Natural draft cooling tower, discharge noise	—	—	100	99	101	103	105	107	105	112

Motors between 750 and 4000 kW[1] The sound power level for large drip-proof electric motors (between 750 and 4000 kW) can be estimated by using Table 2.

2.3 Gas Turbines[2]

Manufacturers of gas turbines will likely have sound power level data for the exhaust and inlet. Obtaining accurate casing data can be difficult due to contributions from the inlet, exhaust, or other equipment. The overall sound power level for gas turbines can be estimated using the following expression:

$$L_w = 127 + 15 \log \mathrm{MW} \quad \text{(intake)}, \tag{7}$$

$$L_w = 133 + 10 \log \mathrm{MW} \quad \text{(exhaust)}, \tag{8}$$

$$L_w = 120 + 5 \log \mathrm{MW} \quad \text{(casing)}, \tag{9}$$

where MW is the power rating of the turbine. The octave band sound power level for each of these sources can be obtained by subtracting the values shown in Table 1.

2.4 Reciprocating Engines[2]

The major noise sources of natural-gas and diesel reciprocating engines are the engine casing, exhaust noise, and for turbocharged engines, the air inlet.

Engine Casing Noise The sound power level for engine casing noise can be calculated using the equation

$$L_w = 94 + 10 \log \mathrm{kW} + A + B + C + D, \tag{10}$$

where L_w is the overall sound power level of the casing, kW is the full-load rating of the engine, and A, B, C, and D are correction factors obtained from Table 3. The octave band sound power levels can be obtained by subtracting the values shown in Table 1.

Air Inlet For those engines with a turbocharger, the air inlet noise can be calculated using the equation

$$L_w = 95 + 5 \log \mathrm{kW} - \frac{l_{\mathrm{inlet}}}{1.8}, \tag{11}$$

where L_w is the overall sound power level of the air inlet, kW is the full-load rating of the engine, and l_{inlet} is the length of the inlet ducting in metres. For engines under about 340 kW, the calculation for the casing noise also takes into account the air inlet noise for both turbocharged and naturally aspirated engines.

The octave band sound power levels can be obtained by subtracting the values shown in Table 1.

Engine Exhaust The sound power level for the unmuffled exhaust can be calculated using the equation.

$$L_w = 120 + 10 \log \mathrm{kW} - T - \frac{l_{\mathrm{exhaust}}}{1.2}, \tag{12}$$

where L_w is the overall sound power level of the exhaust, kW is the full-load rating of the engine, T is a correction term for engines with turbochargers ($T = 6$ if the engine has a turbocharger, $T = 0$ if there is no turbocharger), and l_{exhaust} is the length of the exhaust in metres. The octave band sound power levels can be obtained by subtracting the values shown in Table 1.

2.5 Steam Turbines[1]

Manufacturers of steam turbines will likely have sound pressure level measurements, but they may not have been

TABLE 3 Sound Power Level Correction Terms for Casing Noise of Reciprocating Engines

Correction Term	Correction Type	Qualifying Condition	dB
A	Speed	Less than 600 rpm	−5
		600–1500 rpm	−2
		Over 1500 rpm	0
B	Fuel	Diesel only	0
		Diesel and/or natural gas	0
		Natural gas only (with small amount of "pilot oil")	−3
C	Cylinder arrangement	In-line	0
		V-type	−1
		Radial	−1
D	Air intake	Unducted air inlet to unmuffled Roots blower	+3
		Ducted air from outside the engine compartment enclosure	0
		Ducted air to muffled Roots blower	0
		All other types of inlets (with or without a turbocharger)	0

measured in accordance with a recognized standard. The overall sound power level for steam turbines can be estimated using the expression

$$L_w = 93 + 4\log \text{kW}. \qquad (13)$$

The octave band sound power levels can be obtained by subtracting the values shown in Table 1.

2.6 Steam Turbine–Generator Units[1]

The sound power level for main steam turbine–generator units (between 200 and 1100 MWe) can be calculated using the equation

$$L_w = 113 + 4\log \text{MWe}, \qquad (14)$$

The octave band sound power levels can be obtained by subtracting the values shown in Table 1. Equation (14) takes into account sound contributions from the high- and low-pressure turbines, generators, and shaft-driven exciters. Noisy couplings and steam control valves can cause higher sound power levels than those predicted with this equation.

2.7 Transformers[1]

Noise from the body of the transformer is made up of tones at the even harmonics of the line frequency (120, 240, 360, 480, ... Hz) with the 120-Hz tone being the dominant sound at normal receiving distances. When additional cooling of the transformer is needed, noise from the cooling fans can become the dominant noise source. Lower fan speeds and optimal blade shapes have helped make newer fans much quieter, which is particularly helpful in noise-sensitive applications.

The National Electrical Manufacturers Association (NEMA) sound pressure level rating is the A-weighted sound pressure level one foot from the transformer and can be estimated from the following formulas:

$$\text{NEMA sound rating} = 55 + 12\log \text{MV(A)} \quad \text{for a standard transformer,} \qquad (15)$$

$$\text{NEMA sound rating} = 45 + 12\log \text{MV(A)} \quad \text{for a quieted transformer.} \qquad (16)$$

These equations are valid for transformers between 20 and 450 MV(A). The A-weighted sound power level can be calculated using the formula

$$L_w = \text{NEMA sound rating} + 10\log S, \qquad (17)$$

where L_w is the A-weighted sound power level and S is the surface area of the four side walls in square metres.

The term $10\log S$ may be estimated from the MV(A) rating by using the formula

$$10\log S = 14 + 2.5\log \text{MV(A)}. \qquad (18)$$

Octave band sound power and pressure levels can be obtained by *adding* the appropriate octave band correction factors shown in Table 1.

3 DRIVEN EQUIPMENT

3.1 Air Compressors

Centrifugal Compressors[3] The sound power level radiated from the casing or discharge piping for large centrifugal compressors can be calculated using the equation

$$L_w = 3 + 20 \log \text{kW} + 50 \log U - 17 \log(mf), \qquad (19)$$

where U is impeller tip speed in metres per second ($30 < U < 230$), m is the surface weight of the casing or pipe wall in kilograms per square metre, and f is the octave band center frequency.

The frequency of the maximum sound power is given as

$$f_p = 4.1U. \qquad (20)$$

For the octave band containing f_p, the sound power level is the value from Eq. (19) minus 4.5 dB. The adjacent octave bands above and below f_p should roll off at a rate of 3 dB/octave.

Reciprocating Compressors[3] The octave band sound power levels radiated from the casing or discharge piping for large reciprocating compressors can be calculated using the equation

$$L_w = 154.5 + 10 \log \text{kW} - 17 \log(mf), \qquad (21)$$

where kW is the power of the driver motor, m is the surface weight of the casing or pipe wall in kilograms per square metre, and f is the octave band center frequency.

This equation assumes that the sound power is radiated from 15 m of discharge piping. To determine the octave band levels, first calculate the fundamental frequency [$f_p = B$ (rpm)/60, where B is the number of cylinders]. For the octave band containing f_p, the sound power level is the value from the equation minus 4.5 dB. The adjacent octave bands above and below f_p should roll off at a rate of 3 dB/octave.

If only the rated power of the compressor is known, the following may be used to calculate the sound power level of centrifugal, rotary, and reciprocating compressors.

Centrifugal Air Compressors[1] The sound power level for the casing noise of centrifugal compressors between 1100 and 3700 kW can be calculated using the equation

$$L_w = 79 + 10 \log \text{kW}, \qquad (22)$$

where L_w is the overall sound power level of the casing and kW is the rated power of the compressor. The octave band sound power levels can be obtained by subtracting the values shown in Table 1.

The sound power level for the unmuffled air inlet of centrifugal compressors can be calculated using the equation

$$L_w = 80 + 10 \log \text{kW}. \qquad (23)$$

The octave band sound power levels can be obtained by subtracting the values shown in Table 1.

Rotary and Reciprocating Air Compressors[1] The sound power level for rotary and reciprocating air compressors, including partially muffled air inlets, can be calculated using the equation

$$L_w = 90 + 10 \log \text{kW} \qquad (24)$$

where kW is the rated power of the compressor. The octave band sound power levels can be obtained by subtracting the values shown in Table 1.

3.2 Boiler and Reactor Feed Pumps[1]

The sound power level for large boiler and reactor feed pumps driven by either motors or steam turbines can be estimated by using Table 4.

The larger pumps, which are generally driven by steam turbines, have high tonal noise components that increase the noise levels by several decibels in the

TABLE 4 Sound Power Levels of Boiler and Reactor Feed Pumps

Pump Power Rating (kW)	Sound Power Level	
	Linear	A-Weighted
1,000	108	104
2,000	110	106
4,000	112	108
6,000	113	109
9,000	115	111
9,500	113	112
12,000	115	114
15,000	119	118
18,000	123	122

250–2000-Hz octave bands. Thus, the sound power ratings in Table 4 rise sharply for pumps with high power ratings.

The octave band sound power levels can be obtained for two ranges of power ratings by subtracting the values shown in Table 1.

3.3 Fans[1]

Centrifugal-Type Fans with Airfoil or Backward-Curved Blades The sound power level of the inlet noise of forced-draft fans with inlet control vanes or the discharge noise of induced-draft fans can be calculated using the equation

$$L_w = 10 + 10\log Q + 20\log \delta P, \qquad (25)$$

where Q is the air flow in cubic metres per minute and δP is the pressure drop across the fan in newtons per square metre. The octave band sound power levels can be obtained by subtracting the values shown in Table 1.

Five decibels should be added in the octave band containing the blade passage frequency for forced-draft fans used in high-pressure applications. For induced-draft fans, add 10 dB in the bands containing the blade passage frequency and its second harmonic. This correction term is called the blade frequency increment. The blade passage frequency is calculated as follows:

$$\text{Blade passage frequency} = (\text{fan rpm} \times \text{number of blades})/60. \qquad (26)$$

Centrifugal-Type Fan Casing If there is ducting on the inlet to the fan, the overall sound power level of the uninsulated fan casing may be calculated using the equation

$$L_w = 1 + 10\log Q + 20\log \delta P. \qquad (27)$$

The octave band sound power levels can be obtained by subtracting the values shown in Table 1.

Add 5 dB in the octave band containing the blade passage frequency for high-pressure applications.

The noise generated at the uninsulated fan breaching (discharge ductwork) is approximately 6 dB lower than than that generated by the uninsulated fan housing. The frequency spectrum is similar to that for the uninsulated fan housing.

Gas Recirculation Fan Casing The overall casing sound power level for centrifugal fans used for gas recirculation service (1300–4100 kW) can be calculated using the equation

$$L_w = 13 + 10\log Q + 20\log \delta P. \qquad (28)$$

The octave band sound power levels can be obtained by subtracting the values shown in Table 1.

Axial-Flow Fans The overall sound power level of the inlet noise of forced-draft fans or the discharge noise of induced-draft fans can be calculated using the equation

$$L_w = 24 + 10\log Q + 20\log \delta P. \qquad (29)$$

The octave band sound power levels can be obtained by subtracting the values shown in Table 1, except for the octave band containing the blade passage frequency, where 6 dB should be added rather than subtracting the value in the table. For vane axial fans, inlet guide vanes cause much higher blade passage noise than outlet guide vanes.

3.4 Gears[2]

The following equation provides an estimate of gearbox noise levels based on studies of gearboxes rated from 200 to 17,500 kW:

$$L_w = 79 + 3\log \text{rpm} + 4\log \text{kW} + 10\log A, \qquad (30)$$

where L_w is the octave band sound power level, rpm is the speed of the slower gear shaft, kW is the power transmitted through the gearbox, and A is the conformal surface area (in square metres) at 1 m from the gearbox.

Since the octave bands containing the gear-meshing and ringing frequencies are different according to the particular gearbox being used, the level predicted by the equation is an estimate for all bands equal to or greater than 125 Hz. For the 31-Hz band subtract 6 dB, and for the 63 Hz band, subtract 3 dB from the level obtained from Eq. (30).

3.5 Generators[3]

The sound power level for generators (not including the driver noise) can be calculated from the equation

$$L_w = 84 + 10\log \text{MW} + 6.6\log \text{rpm}, \qquad (31)$$

where L_w is the overall sound power level, MW is the power rating of the generator, and rpm is the speed of the generator. The octave band sound power levels can be obtained by subtracting the values shown in Table 1.

3.6 Pumps[2]

The sound power level for motor-driven pumps under 2000 kW can be calculated using Table 5. In this table, kW is the nameplate motor rating and A is the conformal surface area (in square metres) at 1 m from the pump. The octave band sound power levels can be obtained by subtracting the values shown in Table 1. These equations are based on sound pressure level measurements at 1 m from the pump and therefore include some sound contributions from the driver and suction and discharge piping.

For high-pressure applications, the level in the band that contains the blade passage frequency and its second harmonic may be 5–10 dB higher than the calculated levels.

4 VALVE AND PIPING NOISE

The sound level produced by gas control valves is dependent on whether the flow through the valve is subcritical or supercritical (choked-valve condition). For subcritical flow, the valve noise is due to turbulent mixing and turbulence–boundary interaction. For supercritical flow, the valve noise is due primarily to broadband shock noise. Many valve manufacturers, such as Fisher or Masoneilan, have semiempirical models to predict the sound pressure level at a position 1 m downstream of the valve and 1 m out from the piping. These models are expected to predict the A-weighted sound level to within 5 dB of the actual level. For situations when the manufacturer cannot provide data or as an alternative, Ng[4] provides a control valve calculation procedure. All of these methods assume some number of straight piping sections downstream of the valve. The presence of a tee, mitered bend, or an elbow near the valve discharge can significantly increase the sound level from quiet trim valves.

Turbulence generated in piping can also be a significant source of noise. Concowe[5,6] and Norton[7] have published procedures by which flow-induced sound power levels can be predicted. These procedures require a fairly detailed knowledge of the flow conditions and piping and are beyond the scope of this chapter but are expected to predict the A-weighted sound level to within 3 dB.

5 INDUSTRY-SPECIFIC EQUIPMENT

5.1 Air Conditioning for Buildings

Cooling Towers[1]

Mechanical-Draft Cooling Towers The sound power level of mechanical-draft propeller-type cooling towers can be calculated using the equations

$$L_w = 96 + 10 \log \text{kW} \quad \text{(for full-speed operation)}, \tag{32}$$

$$L_w = 88 + 10 \log \text{kW} \quad \text{(for half-speed operation)}, \tag{33}$$

where L_w is the overall sound power level and kW is the full-speed power rating of the fan. The octave band sound power levels can be obtained by subtracting the values shown in Table 1.

When calculating sound pressure levels at distances greater than 10–20 m from the tower, *add* the values shown in Table 6 to account for directional effects.

Natural-Draft Cooling Towers The sound power level of the rim, located at the base, and the discharge, located at the top, of large natural-draft cooling towers can be estimated by using the values shown in Table 2. Since the low-frequency noise levels of the water flow are not significant, sound power levels are not listed for the 31.5- and 63-Hz octave bands.

In the horizontal direction and 30° above and below the horizontal direction, the noise from the rim may be assumed to radiate uniformly. Most of the noise from the discharge is radiated upward. When calculating sound pressure levels from the discharge, *add* the values shown in Table 6 to account for directional effects.

TABLE 5 Overall Sound Power Level of Motor-Driven Pumps

Operating Speed (rpm)	For Motor Ratings under 75 kW	For Motor Ratings above 75 kW
450–900	$68 + 10 \log \text{kW} + 10 \log A$	$81 + 3 \log \text{kW} + 10 \log A$
1000–1500	$70 + 10 \log \text{kW} + 10 \log A$	$83 + 3 \log \text{kW} + 10 \log A$
1600–1800	$75 + 10 \log \text{kW} + 10 \log A$	$88 + 3 \log \text{kW} + 10 \log A$
3000–3600	$72 + 10 \log \text{kW} + 10 \log A$	$85 + 3 \log \text{kW} + 10 \log A$

TABLE 6 Octave Band and A-Weighted Sound Power Level Adjustments for Directional Effects of Cooling Towers

Source	31.5	63	125	250	500	1000	2000	4000	8000	*A*
Mechanical draft, air inlet side	0	0	0	1	2	2	2	3	3	2
Mechanical draft, enclosed side	−3	−3	−3	−3	−3	−3	−4	−5	−6	−4
Mechanical draft, top	3	3	3	3	3	4	4	3	3	3
Natural draft, top	8	8	8	8	9	9	10	10	10	10
Natural draft, horizontal	−2	−3	−4	−6	−8	−10	−12	−14	−16	11
Natural draft, 30° below horizontal	−3	−4	−5	−7	−9	−12	−14	−16	−18	13

Packaged Chillers with Compressors

Packaged Chillers with Reciprocating Compressors[8] The A-weighted sound power level can be calculated from the equation below as a function of tons of refrigeration (1 ton of cooling capacity is equal to 3.52 kW of heat removal). The standard error of this equation is estimated to be 5 dB. The noise levels are due primarily to the compressor component and not the drive motor:

$$L_w = 71 + 9\log(\text{tons of cooling capacity}) + 10\log A, \tag{34}$$

where A is the conformal surface area (in square metres) at 1 m from the unit. The octave band sound power levels are obtained by subtracting the values shown in Table 1.

Packaged Chillers with Rotary-Screw Compressors[2] The overall sound power level of rotary-screw compressors between 100 and 300 tons of cooling capacity and operating at approximately 3600 rpm is given by the equation

$$L_w = 95 + 10\log A. \tag{35}$$

The octave band sound power levels can be obtained by subtracting the values shown in Table 1.

Packaged Chillers with Centrifugal Compressors[8] The A-weighted sound power level can be calculated from the equation below. The standard error of this equation is estimated to be 4 dB. For larger units with built-up assemblies, use the conformal area of the compressor section only:

$$L_w = 60 + 11\log(\text{tons of cooling capacity}) + 10\log A. \tag{36}$$

The octave band sound power levels can be obtained by subtracting the values shown in Table 1.

5.2 Construction Equipment[1]: Diesel-Powered, Mobile Equipment

The sound power level for diesel powered, mobile equipment such as crawler tractors, dozers, tractor shovels, front-end loaders, backhoes, graders, mobile cranes, and trucks can be calculated using the equation

$$L_w = 99 + 10\log \text{kW}, \tag{37}$$

where L_w is the overall sound power level and kW is the power rating of the engine. This equation is for turbocharged or naturally aspirated engines with conventional exhaust mufflers. In the typical application, the sound level will average about 4 dB lower than the calculated level since the engine is frequently not operated in the maximum-power condition. The octave band sound power levels can be obtained by subtracting the values shown in Table 1.

6 OIL FIELD EQUIPMENT

Increased governmental noise regulations and concerns about drilling in environmentally sensitive locations have created a need for drilling rig sound level analyses. Example sound power level data for a standard degasser, drawworks, mud pump, and top drive are presented in Table 2. Some manufacturers offer special noise control options that can provide a reduction of 5–10 dB(A) from the values provided in the table. The sound power for the degasser is typical for a unit rated at 227 m^3/h and includes the contribution from the drive motor. The sound power levels for the mud pump and drawworks do not include the drive motors.

APPENDIX: CALCULATION OF CONFORMAL SURFACE AREA

A conformal surface is a hypothetical surface located a distance d from the nearest point on the envelope of the reference box. It is different from a rectangular surface because a conformal surface has rounded corners. For small surfaces, the difference in area between a rectangular surface and a conformal surface is typically small ($10 \log A < 1\text{dB}$); however, for large surfaces the difference in areas can be significant ($10 \log A > 2$ dB). The area of a conformal surface can be calculated from the equation

$$S = 4(ab + bc + ca)\,\frac{a + b + c}{a + b + c + 2d},$$

where $a = 0.5L + d$, $b = 0.5W + d$, and $c = H + d$, with d the measurement distance, L the length of the reference box, W the width of the reference box, and H the height of the reference box.

REFERENCES

1. L. N. Miller, E. W. Wood, R. M. Hoover, A. R. Thompson, and S. L. Patterson, *Electric Power Plant Environmental Noise Guide*, rev. ed., Edison Electric Institute, Washington, D.C. 1984, Chapter 4.
2. *Noise and Vibration Control for Mechanical Equipment*, Manual TM5-805-4/AFM 88-37/NAVFAC DM-3.10, manual prepared by Boit Beranek and Newman for the Joint Department of the Army, Air Force, and Navy, Washington, DC, 1980, Chapter 7.
3. D. A. Bies and C. H. Hansen, *Engineering Noise Control*, Unwin Hyman, London, 1988, Chapter 11.
4. K. W. Ng, "Control Valve Aerodynamic Noise Generation and Prediction," *Proceedings of NOISEXPO*, *80, Acoustical Publications, Inc., Bay Village, OH, 1980, pp. 49–54.
5. J. N. Pinder, "The Study of Noise From Steel Pipelines," CONCAWE Report No. 84/55, 1984, Brussels, Belgium.
6. P. H. M. Corbet and P. J. van de Loo, "Experimental Verification of the ISVR Pipe Noise Model," CONCAWE Report No. 84/64, 1984, Brussels, Belgium.
7. M. P. Norton and A. Pruiti, "Universal Prediction Schemes for Estimating Flow-Induced Industrial Pipeline Noise and Vibration," *Appl. Acoust.*, Vol. 33, 1991, pp. 313–336.
8. ASHRAE, "Sound and Vibration Control," in *HVAC Applications*, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., Atlanta, 1991, Chap. 42.

87

AIRPORT NOISE

KENNETH MCK. ELDRED

1 INTRODUCTION

The noise associated with civil aircraft operations in the vicinity of airports has been the major environmental noise problem in the United States and in many other countries since the late 1950s, when turbojet aircraft entered service. The public opposition to increases in aircraft noise is the major factor that has prevented most development of new airports. This problem is worldwide, and the environmental restrictions on airport development are found in Japan, Germany, England, and many other countries.

The civil aviation jet age in the United States opened with a roar in 1958, with the beginning of transcontinental and transoceanic flight operations by Boeing 707 and Douglas DC-8 turbojet aircraft. Their takeoff noise, expressed in maximum A-weighted sound level, was about 20 dB greater than that of the propeller aircraft that they replaced. The Port Authority of New York promulgated its takeoff noise regulation with limits of 112 Pndb[1] at locations at 2.7–4.2 statute miles from the start of takeoff roll with 24-h noise monitoring at several locations to determine compliance. London Heathrow, with its jet operations, also including the Comet and Caravelle, imposed night and noise level restrictions. Soon other major airports developed rules to attempt to control the noise.[2]

The noise of the early jets was controlled by a variety of multiple and corrugated nozzle schemes that shifted much of the sound energy to higher frequencies. However, these devices had significant performance penalties of about 1% thrust loss per decibel reduction. Within 2 years, engine manufacturers introduced low bypass turbofan engines by adding a bypass co-annular flow fan stage to the turbojets used in the first generation of civil engines. This development significantly lowered the full power jet noise and added takeoff thrust, which improved performance. Unfortunately, the fans sounded like sirens, resulting in trading the improvement in jet noise during takeoff for a large increase in approach noise. Finally, by the late 1960s, the world's leading engine manufacturers had developed the familiar high-bypass-ratio engines, which were much quieter than the earlier engines, particularly on takeoff, and provided much higher propulsion efficiency. Also, concerted government and industry research led to the development of quieter fans and turbines, providing a significant reduction in the pure tones that were characteristic of the earlier turbofan engines.

This chapter addresses the assessment and control of civil airport noise in the United States. Then, it reviews the principal national and airport-specific actions undertaken to reduce the noise and concludes with a summary of the changes in the estimated aircraft noise population impact in the United States from 1960 to 2000. Although the principal examples and discussion of this topic are oriented to commercial civil aviation in the United States, the general problems and approaches to solutions apply at civil, military, and general aviation airports throughout the world.

2 ASSESSMENT OF AIRPORT NOISE

The assessment of aircraft noise in the vicinity of an airport involves measurement and/or calculation of the noise and estimating its impact, primarily on people who reside in areas that are subject to the noise. In these assessments it is usually necessary to describe the noise

---

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker
Copyright © 1997 John Wiley & Sons, Inc.

of a single aircraft flight (single event sound) and to quantify the cumulative noise of many flights throughout a 24-h day. The reference pressure is 20 μPa. The descriptors for the sound of a single event include the following:[3,4]

Sound Level. Usually level of the ratio of the mean-square A-weighted sound pressure obtained with the slow time constant to a squared reference sound pressure in decibels.

Maximum Sound Level (MXSL). Greatest slow (1-s) sound level within a stated time interval or during a specific single event in decibels.

Perceived Noise Level (PNL). A frequency-weighed sound pressure level that combines the sound pressure levels in 24 one-third-octave bands from 50 Hz to 10 kHz in decibels.[5]

Sound Exposure Level (SEL). Usually the level of the ratio of a given time integral of the squared instantaneous A-weighted sound pressure over a stated time interval or event to the product of the squared reference sound pressure of 20 μPa and reference duration of 1 s in decibels.

Effective Perceived Noise Level (EPNL). Level of the time integral of antilogarithm of one-tenth of the tone corrected perceived noise level over the duration of an aircraft flyover and with a reference duration of 10 s in decibels.[5]

The descriptors for the cumulative noise of many events include the following:

Equivalent Sound Level (LEQ). Usually the level of the ratio of the time average integral of the squared instantaneous A-weighted sound pressure during a stated time interval (often 1 h or 1 day) to the product of the squared reference pressure and the duration of the time interval in seconds in decibels.

Day/Night Sound Level (DNL). Usually the level of the ratio of the time average integral of the squared instantaneous A-weighted sound pressure during a 24-h period after multiplying by 10 the values from 10 p.m. to 7 a.m. to the product of the squared reference pressure and 86,400 s in decibels.

Community Noise Equivalent Level (CNEL). Usually the level of the ratio of the time average integral of the squared instantaneous A-weighted sound pressure during a 24-h period after multiplying by 5 the values between 7 p.m. and 10 p.m. and multiplying by 10 the values between 10 p.m. and 7 a.m. to the product of the squared reference pressure times 86,400 s in decibels. *Note:* The CNEL is used in California instead of the DNL.[6]

The primary descriptor used in the United States for airport noise is the annual average value of the DNL. For airport projects that are funded partially or wholly by the federal government, the environmental assessment procedures of the Federal Aviation Administration (FAA) require use of the DNL, supplemented by other descriptors, where necessary. The magnitude of impact is generally measured by counting the number of people estimated to reside in areas where the DNL from airport operations exceeds 65 dB (for some projects 60 dB) and extending to the highest value of DNL, usually in intervals of 5 dB.

2.1 Monitoring Airport Noise

Airport noise may be monitored by several methods, including continuous or intermittent sound measurements, with or without an observer present, combined sound measurement with radar flight track data, or estimates based on calculations using aircraft operations data in a computer model.

Permanently installed and/or portable field microphone systems are used to monitor aircraft noise at many airports. The first permanent system was installed in 1959 by the Port Authority of New York and New Jersey.[7] But it was not until the early 1970's, when the State of California promulgated an Airport Noise Standard,[6] that permanent noise-monitoring systems began to be installed at many other airports in the United States. The California Standard addressed both permanent and portable noise measurements. For the latter, it specified that measurements be made for 4 weeks during the year, 1 week during each season. The California standard became a de facto specification for airport noise-monitoring systems.

Typical airport noise-monitoring systems have from 5 to 25 remote sound monitors that provide continuous 24-h coverage. The remote monitors are usually connected to the central system by standard telephone lines and transmit data either continuously or when dialed by the central system for a data download. Some systems also have a direct telephone line that may be used to listen to the noise. Most systems can provide noise samples at intervals of 1 s or less. They also can provide the maximum sound levels and sound exposure levels for sound events that meet preset criteria (e.g., sound level thresholds, duration, rise and fall time). These sound event criteria are usually set to discriminate against nonaircraft sound events so that the reported measured sound events have a high probability to be aircraft related. Concurrently, or at a later time, the true sources for the sound events can be determined by the system's operator or automatically by a computer system by matching other data on the time of occurrence, runways in use, flight track data, if available, and identity of aircraft arrivals and departures. Most systems provide the hourly equivalent sound level and the day–night sound level, sometimes subdivided into aircraft and nonaircraft noise components. Often, other quantities are available such as the statistical sound levels exceeded for stated percentage of time, the maximum sound level, and the number of sound events in hourly and 24-h periods.

The locations selected for remote monitors depend on the purpose of the measurements. For airports in California, many of the monitors are located in the vicinity of the noise impact boundary, which approximates the CNEL of 65 dB noise contour in areas where the land use is residential or in other noise-sensitive uses. Often sound monitors are located either directly under flight paths or to each side of the expected ground track to bracket the flight paths. Sound monitors used to check aircraft adherence to noise abatement procedures often are located in the vicinity of the nearest residences affected by noise from operations on the flight path to be monitored. They are also used at greater distances from the airport to monitor flight path procedures that incorporate turns to avoid flying near to specified residential areas or to determine the sounds that cause complaints about aircraft noise in distant communities. In some cases monitors are located under the flight path at specific distances from the start of roll for departures or landing threshold for arrivals to compare measured values with noise limits in an airport's noise regulations.

The data gathered from the sound monitors may be used for many purposes, including monitoring the cumulative noise and its change with time, the maximum noise at specific locations with respect to noise regulations, and the noise of specific flights by aircraft type, airline, and so on, to compare the relative effectiveness of flight procedures for noise abatement. At many airports selected summary data are published on an annual basis to keep the public informed of the quantitative progress of the airport's noise control program.* The availability of factual data is often of significant importance in providing a basis for substantive discussion of noise issues.

2.2 Monitoring Fight Paths

At the majority of airports in the United States used for commercial turbojet aircraft, the FAA uses an Aircraft Radar Tracking System (ARTS) for real-time monitoring of aircraft position so that air traffic controllers can see the relative positions of all aircraft in their assigned control sectors. The system (currently ARTS-3 at most airports) determines the *X*, *Y*-position of the aircraft by radar echoes at intervals of about 4 s. It determines the height and identity of each aircraft by the radio return data from a transponder on the aircraft that is activated by the radar signal. Each aircraft is assigned a transponder code for a specific flight plan, so that the controller radar display can give the height as well as the identity of the aircraft (airline flight number and type) and other data from the flight plan associated with the transponder code.

These radar data are retained for a limited time by the FAA on digital disk or tape and are accessed by some airports[8] to obtain flight path data and aircraft identity after a 15-day waiting period. In many other countries the airport radar is available in real time to a system that correlates flight track data with sound event data measured at remote monitors. A few airports obtain aircraft path data from other passive systems that use the radar and its triggered transponder to independently calculate path data. These passive systems operate in real time but cannot directly obtain an aircraft's identity, only its transponder code.

Radar-derived flight path data may be used to determine the sources of measured sound events, providing a greater certainty of the aircraft flight-related events. Further, they may be used to evaluate the effectiveness of flight track rules and control procedures and for the definition of representative noise paths for computer model analysis.

2.3 Calculating Airport Noise and Population Impact

The calculation of airport noise[10] and potential population impact must account for the major factors that control airport noise impact given in Table 1. Contours of cumulative noise in DNL, or other metrics, for an average day during a specified time period may be computed by the use of the FAA integrated noise model (INM)[10] or the Air Force noise map model.[11] See example in Fig. 1.[12] These models can also be utilized to generate a sound exposure level (SEL) contour for a single flight of a specific type of aircraft on a specific or generalized flight track.

These computer programs contain an aircraft noise data base specifying for each aircraft type the relationship between sound exposure level, slant distance from receiver to the aircraft, and engine thrust setting. In addition, the INM contains a set of flight profiles for each aircraft type that give the height, engine thrust, and speed for takeoffs at various aircraft weights as a function of distance from the start of takeoff roll and a standard pro-

---

*The California regulation requires quarterly submittals to the state from airports that are required to monitor noise. These records are one of the largest public sources of data regarding the effect of weather and operational changes on the propagation and generation of aircraft noise.

TABLE 1 Principal Factors That Control Airport Noise Impact

Aircraft Noise Characteristic
Noise vs. distance by aircraft type
Number of operations by aircraft type
Ratio of the number of daytime to nighttime operations by aircraft type
Airport Noise Impact Potential
Flight procedures (throttle and flap management) used for landing and takeoff by aircraft type
Stage lengths (takeoff and landing weights by aircraft type)
Spatial configuration and relative utilization of the flight tracks by aircraft type
Spatial distribution of population

file for landing. The input data for each specific case includes the runway and flight track layout; the number of takeoff and landing operations in day and night, distance to departure destination, and runway/track utilization, all by aircraft type; and supplemental data for special situations.

The noise contours can be superimposed over maps containing land use or population data, enabling determination of the land areas with specified uses or the residential population within an interval between any two noise contours. An alternative method for estimating the population residing at various noise levels is to calcu-

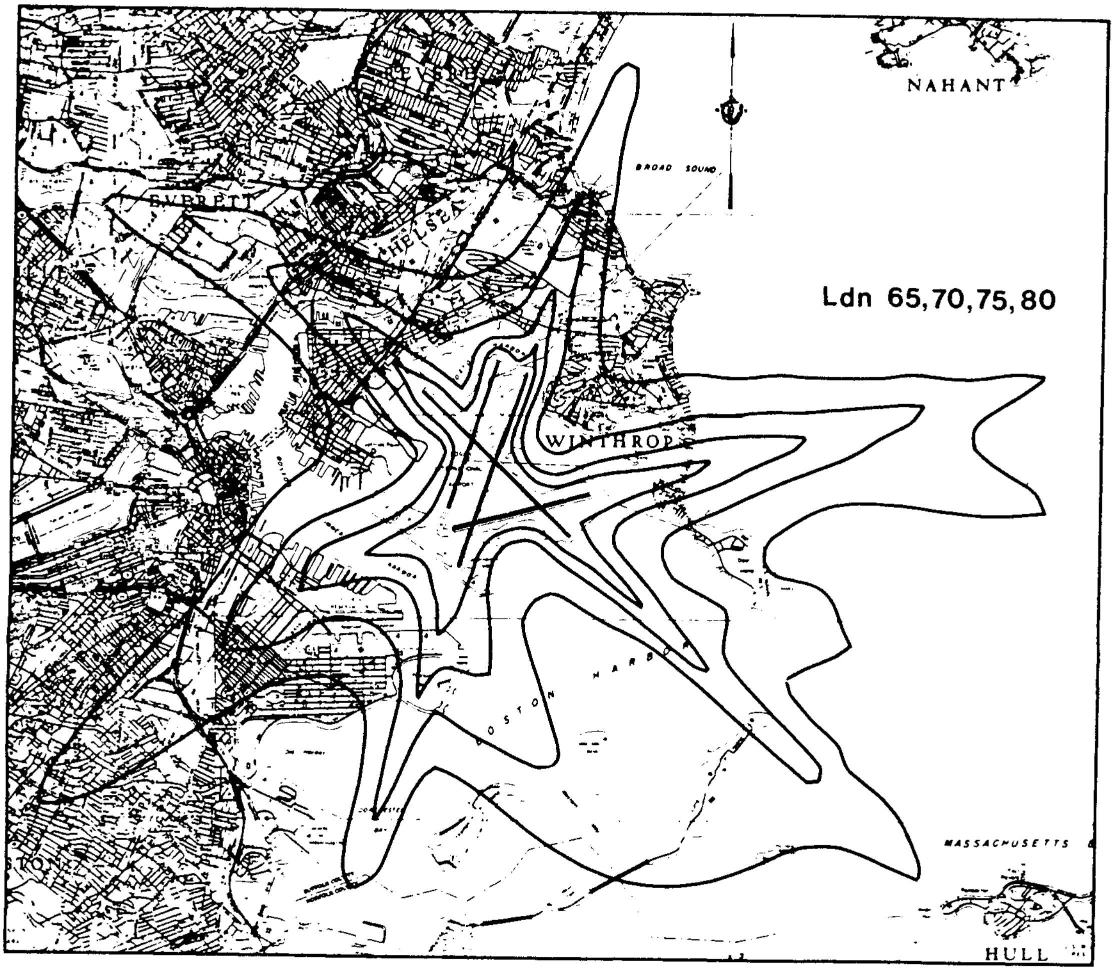


Fig. 1 Example of DNL noise contours at Boston Logan International Airport for 1980 operations.[12]

late the noise level at the centroids of population census blocks or block groups. The results can be summarized in appropriate intervals for display and may be used to ascertain the estimated change in noise at each population centroid for various study alternatives.

The calculations may be used to supplement the measured noise exposure data and to provide insight on the causes of the noise at various locations. The measured data may be used to validate the calculations that usually are within ±2 dB for cumulative noise level if the measurement time period is sufficiently long for the types of variation encountered at the airport; if the operation is accurately defined as to aircraft type, stage length, time of day, and runway use; and if the flight tracks are modeled in sufficient detail. Once a model of the noise at an airport is verified and accepted by the local authority, it becomes the basis for studying the potential effects of proposed future noise abatement actions.

3 NOISE ABATEMENT AT THE NATIONAL LEVEL

The U.S. national effort to reduce the noise around airports is led by the FAA. It has a comprehensive set of programs including regulations pertaining to noise limits for aircraft types, time schedules for phasing older, noisy aircraft out of the fleet, detailed environmental assessment procedures, and the funding and management of airport specific projects for the control of noise. These programs are addressed in the following section.

3.1 Aircraft-Type Certification for Noise and Phaseout of Noisy Aircraft

In 1958 The Port Authority of New York promulgated a noise rule[1] that contained a perceived noise level limit of 112 dB, applicable to the noise received at specific locations from a departing aircraft. The four current controlling locations at John F. Kennedy International Airport are at an average distance of 5900 m (4770–6690-m range) from the beginning end of the runway along the extension of its centerline. This noise level limit was difficult to meet on hot summer days by the early intercontinental models of the Boeing 707 and Douglas DC-8 aircraft powered by turbojet engines, often requiring a reduction of payload. This rule provided the de facto design requirement for departure noise until 1969, when the FAA promulgated its aircraft-type certification requirements for noise in FAR Part 36.[13]

In FAR Part 36 the FAA specified three locations for the measurement of noise: takeoff, sideline, and approach. The original locations in FAR Part 36 have evolved with international consensus as follows. The takeoff noise measurement position is at 6500 meters from the start of takeoff roll along the extended centerline, the sideline noise measurement position is at the point of maximum takeoff noise along a line parallel to the runway centerline and displaced by 450 m, and the approach noise measurement position is at 2000 m from the runway threshold along the extended centerline.

The Part 36 noise limits are defined in terms of the EPNL descriptor in units of decibels. The noise is measured for each aircraft type under tightly controlled test conditions that are defined in the regulation or is calculated from previous measurements of an earlier model of the same basic aircraft. The original FAA limits[13] for subsonic fixed-wing aircraft are now referred to as stage 2 noise limits, and aircraft certified prior to FAR Part 36 are referred to as stage 1 aircraft (no noise limits). The currently applicable FAA limits, referred to as stage 3 noise limits, were promulgated in 1975.[14] Figures 2–4 illustrate the stage 2 and 3 noise limits for the approach, sideline, and takeoff measurement locations using examples of some of the noisiest models certified. Note that stage 1 aircraft over 75,000 lb maximum takeoff weight have not been allowed to operate at U.S. civil airports since 1985.[15]

The FAR Part 36 regulations have made major improvements in the noise radiated from civil aircraft. Figure 2 shows the stage 2 and 3 noise limits for arrival noise at a location under a 3° sloped flight path and located 2000 m from the landing threshold. It also shows 79 certification data points[15] for selected heavier models of 31 basic aircraft types, representative of each of the three noise stages. It shows that stage 2 put a ceiling on arrival EPNL with reductions up to 13 dB. The stage 3 limit was 4 dB lower than stage 2 and the mean of the stage 3 data is about 4 dB lower than the limit.

Figure 3 contains a similar presentation of the maximum EPNL during takeoff obtained along the sideline located about 450 m from the runway centerline. This maximum level is usually found after the aircraft has climbed to several hundred feet above the ground and is operating with *full takeoff power* and flaps. The stage 3 "mean modern > 75K" line for the aircraft with high-bypass turbofan engines is about 5 dB below the stage 3 limit and 13 dB below the stage 2 limit. One aircraft, the 355,000-lb stage 3 DC8-73, is 16 dB below the stage 1 DC8-63 after reengining with high-bypass-ratio turbofan engines.

However, there are six stage 3 aircraft weighing between 150,000 and 200,000 lb that are just at the noise limit and two aircraft weighing between 700,000 and 750,000 lb that are about 1 dB above the noise limit.*

---

*FAR Part 36 allows for tracking levels up to 2 dB at one location and 1 dB at a second location for 3 dB under limit at the third location.

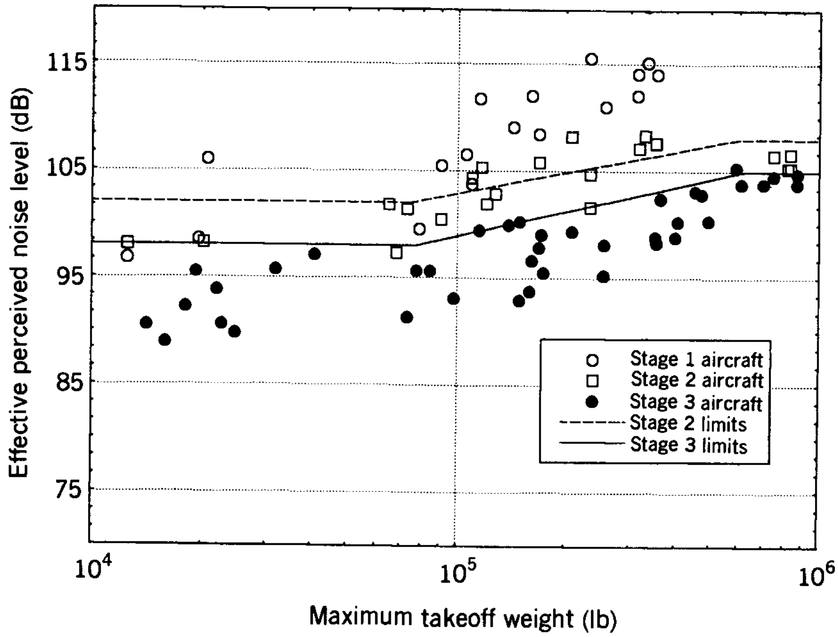


Fig. 2 FAR Part 36 stage 2 and 3 approach noise limits with data for some of the noisier models of various aircraft types.[15]

Four of the six aircraft are Boeing 727-100 and 727-200 with Valsan or Fedex hushkits. The other two aircraft in this group are McDonald Douglas MD-80's. The two heavier aircraft are the Boeing 747-100 and 747-SP recertified to stage 3. Obviously, these hushkitted or recertified aircraft do not represent the same degree of noise reduction technology as do the "mean modern > 75K," which are 5 dB below the limit. Conversely, the

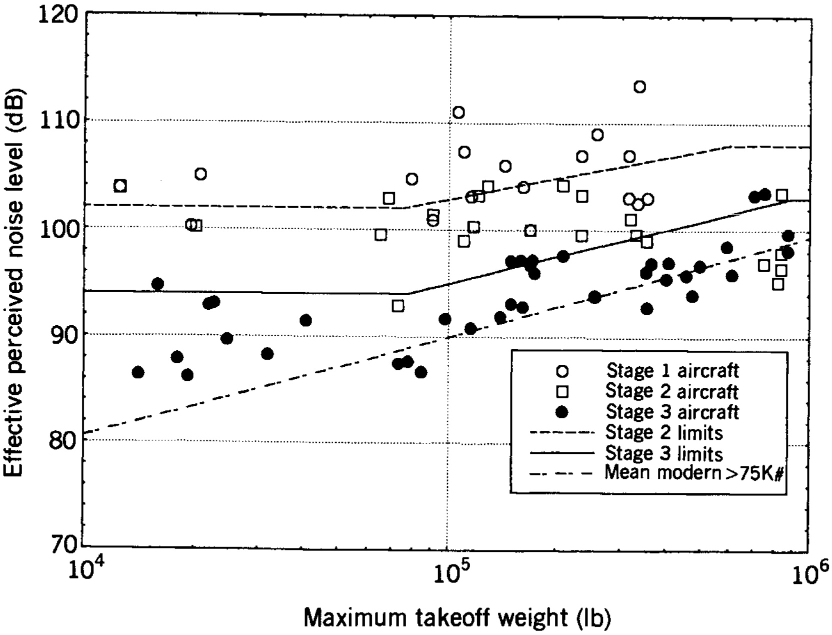


Fig. 3 FAR Part 36 stage 2 and 3 sideline noise limits with data for some of the noisier models of various aircraft types.[15]

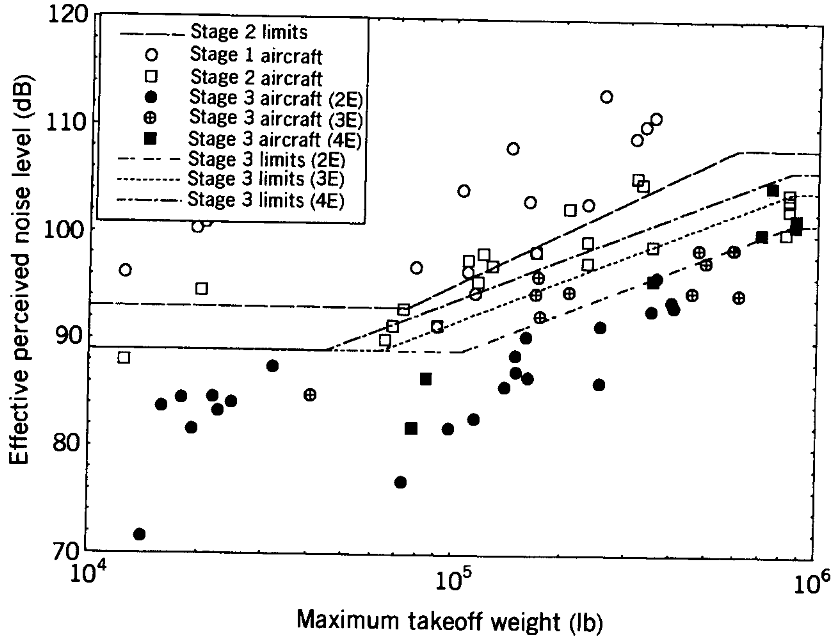


Fig. 4 FAR Part 36 stage 2 and 3 takeoff noise limits with data for some of the noisier models of various aircraft types.[15]

modern engine data demonstrate the technology feasibility of promulgating a rule that lowers the EPNL limits for future aircraft.[16]

Figure 4 is a similar presentation of the maximum EPNL at the takeoff location. The figure is more complex because different stage 3 limits apply to aircraft with two, three, or four engines. These differing limits are the result of the higher climb rates of aircraft with fewer engines, resulting from their higher thrust-to-weight ratio, which in turn is caused by the FAA safety requirements for performance with one engine out. In general, at this location the stage 3 aircraft with modern high-bypass-ratio turbofan engines are still operating at full takeoff thrust and are significantly better than the requirements, whereas the 727 aircraft with low-bypass-ratio JT8D engines and hushkits operate with engines at cutback thrust to just meet the requirements.

Because the stage 3 aircraft with modern turbofan engines have such significant reductions of noise over stage 2 aircraft, many airports have promulgated regulations that restrict operations by stage 2 aircraft during certain hours. The concerns over possible impacts such regulations might have on interstate commerce and the public concern to phase out the noisy stage 2 aircraft led to the passage of the Airport Noise and Capacity Act of 1990.[17] As a result, the FAA has recently amended FAR Part 91[18] to regulate an orderly phaseout of stage 2 aircraft with a maximum certified weight greater than 75,000 lb from the civil fleet by the year 2000. The FAA estimates that this phaseout of stage 2 aircraft will reduce the population living in neighborhoods where the DNL exceeds 65 dB by 85% from the 1990 estimated values.

3.2 Environmental Assessment and Mitigation Policies and Procedures

In response to the requirements of the National Environmental Policy Act of 1969,[19] the FAA established in its Order 1050.1D,[20] a set of policies and procedures for the preparation of environmental impact statements (EISs) and findings of no significant impact (FONSIs) and for preparing and processing environmental assessments (EAs) of FAA actions. For noise, these policies and procedures respond to other legislative requirements, including the Noise Control Act of 1972[21] and the Aviation Safety and Noise Abatement Act of 1979.[22] Additional guidance for preparation and processing the required environmental analyses and statements is contained in the FAA's Airport Environmental Handbook, Order 5050.4A.[23]

In Order 1050.1D the FAA specifies the use of DNL for analysis of the cumulative noise at an airport and the maximum sound level or the sound exposure level for analysis of single-event noise. Further, it prescribes the use of its INM (Noise Map or other FAA approved model) for construction of DNL contours, contains a land use noise compatibility table principally derived from the Federal Inter-Agency Guidelines,[24] and provides additional guidance to the preparer. For evaluation of noise impacts, Order 1050.1D defines the FAA's threshold of "significance" for initial analysis of impacts at specific noise-sensitive areas to be a DNL increase of 1.5 dB when the DNL is greater than 65 dB and a DNL increase of 3 dB when the DNL is 60–65 dB.[25] Proposed actions that result in increases that equal or exceed this threshold require additional analysis as part of the EIS process. However, this threshold does not apply to categorically excluded actions such as changes to air traffic control procedures at altitudes exceeding 3000 ft above ground level.[19]

As a result of these orders, EAs are routinely made on most U.S. airport projects that involve possible noise effects and that are partially or wholly funded by the FAA. Many of these assessments, particularly those related to airport expansion projects, lead to the preparation of EISs. Additionally, many states have similar environmental analysis and reporting requirements.

The FAA has initiated a comprehensive program of airport planning for noise compatibility. This program is codified under FAR Part 150[26] and additional information is provided in FAA Advisory Circular AC 150/5020-1.[27] The performance of this program through 1990 was assessed by the FAA in a report to Congress.[28] The program's objective is to reduce existing noncompatible land uses and to prevent introduction of additional noncompatible uses around an airport. Compatibility is assumed to be achieved when the noise level at a location does not exceed an acceptable value for the land use in that area.

To achieve the program objectives, Part 150 provides the following[26]:

Establishes standard noise methodologies and units
Establishes the INM as the standard computer model
Identifies land uses that normally are compatible or incompatible with various levels of airport noise
Provides for the voluntary development of noise exposure/Maps (NEMs) and noise compatibility programs (NCPs) by airport operations in consultation with all affected parties
Provides for review of NEMs to ensure compliance with Part 150 regulations
Provides for review and approval or disapproval of Part 150 NCPs submitted to the FAA by airport operators

In its approval of NCPs, the FAA has exhibited considerable concern and resistance to proposed airport regulations that restrict airport access. The restrictions of concern include aircraft noise limits at various times of day, limits on operation of aircraft certified under FAR Part 36 stage 2 noise limits, and nighttime curfews. In response to the Airport Noise and Capacity Act of 1990,[17] the FAA issued a new Part 161 entitled Notice and Approval of Airport Noise and Access Restrictions.[29] This regulation contains detailed requirements on the airport operator for providing notice of intent to regulate and stringent requirements for noise and economic impact analyses to be submitted to the FAA for approval.

4 AIRPORT-SPECIFIC NOISE CONTROL MEASURES

Almost all major civil airports in the United States have active noise abatement programs. Most of these utilize the Part 150 process that involves all directly interested parties in the development of consensus solutions. Table 2 summarizes the principal actions that can be taken to abate noise and its perceived problems.

4.1 Optimize Flight Tracks

In developing noise reduction alternatives for a specific airport, one of the first factors to examine is the location of its flight tracks with respect to land use. It is desirable that these tracks be located to make maximum use of noise compatible land such as water, industrial, commercial, and agricultural. Where opportunities exist for changing flight tracks to reduce their noise impact potential, they should be carefully studied with respect to noise benefits, local air space management, fuel, time, cost, and other factors. The results can be dramatic, particularly when large bodies of water are next to the airport.

The Boston Logan runway 22R departure EIS[30] contains one of the more detailed studies of flight track alternatives. Six alternative tracks were studied, comparing potential population impacts for days in which all departures were on runway 22 (220° magnetic). After attaining sufficient altitude, the aircraft turned to one of six headings. These headings varied over 120°, ranging from 220° flying over densely populated urban areas to 100° flying out to sea through the harbor entrance. The populations estimated to be living in areas where the DNL on a maximum day exceeded 65 dB ranged from 91,000 for the base case to 3000 for the harbor entrance alternative that was finally selected.

Although noise analyses are often made with nominal tracks, the tracks of actual flights vary from the nominal track. The amount of this variation for straight tracks may be of the order of ±5°, but the variation is usually much greater for tracks with significant turns. Several airports monitor the actual tracks from the FAA radar data, including the Port Authority of New York and New

TABLE 2 Possible Airport Actions to Abate Noise

Flight tracks	Get aircraft away from people.
Preferential runways	Increase use of runways with least impact.
Restrict noisy aircraft	Minimize operations in day or night.
Noise abatement flight procedures	Require use of noise abatement throttle and flap management procedures for takeoff and/or approach.
Airport layout	Extend or build new runways and taxiways to make best use of compatible land and water.
Shielding barriers	Shield people from noise of ground operations.
Soundproof	Soundproof schools, homes, and churches.
Land use control	Assure compatible land use through acquisition of property or other rights.
Monitor and model	Monitor airport noise and flight tracks to provide data to the public and for evaluating proposed alternatives.
Communication	Listen to complaints and suggestions; develop and institutionalize continuing effective dialogue and information transfer among all concerned parties.

Jersey, which has a system to provide continuous routine track monitoring at its three airports.[8] In most of these situations, defined flight tracks are a principal noise abatement measure.

A variant of the natural flight track dispersion is the use of fanning, a deliberate controlled program of dispersing the flight tracks. There are a variety of techniques employed. For example, in the mid-1970s at Detroit sequential departures on one runway were asked to execute a 10° right turn, 10° left turn, and "no" turn; and presently for runway 31 at New York's LaGuardia Airport each of three tracks at 10° intervals is used for approximately 1 h during non–peak traffic periods prior to switching to the next track. The practice of fanning appears to be accepted as desirable when the noise associated with each track is below the threshold value of significant annoyance for the exposed population. However, if the resulting noise is greater than this threshold, fanning tends to lead to greater numbers of people impacted.

4.2 Define Preferential Runways

Once the flight tracks have been optimized at a specific airport, opportunities for reduction in the noise-impacted population may exist by optimizing the relative utilization of these tracks. The potentials for utilization of the tracks associated with each runway depend upon the wind and weather conditions, the capacity of each runway (or runway combinations) to meet demand, and runway length. In general, preferential runway assignments can only be made in weather conditions that meet the visibility, wind velocity and direction, and runway conditions established by FAA Order 8400.9.[31] The proportion of time that preferential utilization is available is critically dependent on the limit values for tail and cross winds. Small propeller aircraft are typically certificated to lower cross-wind velocities than are jet aircraft and may restrict preferential utilization below that which would result from the FAA order.[31]

The results of implementing preferential runways can also be dramatic in some situations. For example, the nighttime preferential approach to Los Angeles International Airport is from the west over the ocean, rather than from the east over densely populated urban areas. With this procedure, both departures and approaches overfly the ocean. It essentially eliminates nighttime aircraft approach noise in those urban areas for 90–95% of the year when wind direction and speed permit the procedure. Similarly, the nighttime use of runway 15 for departures and 33 for arrivals at Logan International Airport keeps the aircraft over the harbor, rather than over populations in urban areas, when the traffic is only a few aircraft per hour and the wind direction and speed permit the procedure.

The techniques for the implementation of preferential runway systems vary from describing a simple list of runways in order of preference, as above, to sophisticated systems that attempt to achieve multiple goals. If the variable level of demand throughout the day requires more complex runway configurations to provide adequate capacity, then the runway configurations, rather than single runways, must be listed in order of pure preference. If consensus on the order of runways for pure preference cannot be obtained, then it may be possible to obtain agreement among the concerned parties on annual and short-term goals for runway utilization. Implementation of a runway utilization goal system can be based on the expected weather distribution over a long-term data base period, such as 10 years. Preferential use rules can then be formulated that enable long-term average achievement of these runway utilization goals accounting for the expected variation of demand and the relative capacities of the runway configuration. Such a system may also have algorithms to compensate for variations of the weather in any one year from the long-term average weather to enable a closer approximation of the annual goal in each year. The initial Preferential Runway Advisory System for Logan[32] operated on a microcomputer (with manual backup) to attain consensus annual goals and short-term goals to minimize persistent use of a runway combination in a 3-day period and dwell during a 24-h day.

4.3 Restrict Noisy Aircraft

Any restriction on the use of an airport by noisy aircraft must be nondiscriminatory and it must not constitute an interference with interstate commerce. One approach is to pick a maximum A-weighted sound level applicable to overflights at some reasonable location. For example, Santa Monica, California, has a 95-dB maximum limit and has investigated establishing a set of lower limits by aircraft class. A second approach is to require use of the quietest technology. For example, Boston Logan and San Francisco International airports and many others require that night operations be conducted by aircraft that meet FAR Part 36 stage 3 noise limits. A third approach is to combine a single sound level limit with a stage 3 requirement, so that the maximum noise (i.e., aircraft size) is controlled and the noise control technology requirement is consistent for all smaller aircraft. A fourth approach, used by several airports, is to give noise budgets to each airport user that allow the user to trade off between its number of operations and the noise associated with its various aircraft types. These budgets can be designed to be increasingly restrictive in future years.

A basic difference between the two pure approaches—maximum sound level limit and requirement to be certified to a specific FAR Part 36 stage—is that the former only addresses the noisiest aircraft, presumably the largest, most powerful, and/or oldest types. All other aircraft are permitted to make the same amount of noise. On the other hand, a regulation requiring certification to a specific stage should mean that all aircraft that meet the stage noise limits must have incorporated the same level of noise control technology regardless of size and have met the same criterion for safety, technological practicability, economic reasonableness, and appropriateness to type. Thus, the latter approach tends to minimize the total noise exposure, whereas the single limit generally cannot minimize total exposure. An FAA advisory circular[33] gives estimated sound levels at the FAR Part 36 approach and takeoff certification measurement locations that can be of assistance in choosing a specific level appropriate for the airport and time of day. Noise data for both certified and uncertified aircraft are provided by the FAA.[15,34] The data in any of these FAA advisory circulars can be used to determine which aircraft can meet the noise limits and/or stage requirements.

The nighttime noise penalty of 10 dB in the DNL descriptor means that the noise from 1 nighttime operation is counted the same as is the noise from 10 daytime operations of similar character. Thus, for airports that have nighttime operations (10 p.m. to 7 a.m.), their reduction often appears to have considerable potential to reduce noise impact. Often, considerable progress may be achieved through working with the airport's user-operators to eliminate nighttime flights that have only marginal benefit, rescheduling other operations and providing quieter aircraft for nighttime. This latter result also can be achieved for all nighttime operations by choice of a suitable nighttime noise limit. Generally, a noise limit is preferable to a total curfew because a curfew may impede airport service sufficiently to interfere with interstate commerce and may prove very difficult to remove at a later date when aircraft are sufficiently quiet and more service is desired.

4.4 Use Noise Abatement Flight Procedures

Potential noise reductions can be obtained from the use of improved throttle and aircraft configuration techniques for both approach[35,36] and takeoff.[37] For example, lower flap settings reduce aircraft drag, enabling lower engine power settings and, hence, lower noise levels. If these procedures could be properly tailored to specific airport requirements, the gains could even be greater. However, there is considerable controversy over the notion of changing cockpit management procedures from airport to airport, particularly with respect to pilot workload and safety implications, and there are questions of responsibility for adopting specific procedures for specific aircraft types and runways at individual airports.

4.5 Improve Airport Layout

Changes to airport layouts are usually initiated to improve capacity and/or safety. They are rarely made to improve the noise environment because any proposed change to the airport is usually viewed with suspicion by the noise-impacted public.

However, changes in airport layout by extending existing runways, reorienting runways to new directions, or building new parallel runways can in some situations provide significant noise benefits. For example, extensions may be used to have aircraft at greater heights over populated areas. Reoriented runways may be designed to make better use of existing compatible land (or water) for new flight tracks. New parallel runways may be used to move sideline noise further from existing communities and to make better use of existing compatible land. These and other concepts should be reviewed carefully in designing solutions to airport noise problems for their potential applicability.

4.6 Soundproofing and Noise Barriers

The use of physical noise reduction techniques to increase the attenuation in the path between source and receiver is a standard engineering approach. Its implementation in airport noise problems is usually in the sound insulation of schools and residential dwelling units or the construction of sound barrier walls.

In the early 1990s, major sound insulation programs were under way at many airports, including Atlanta, Boston, Denver, Los Angeles, New York, Providence, San Francisco, and Seattle. Most of the programs involve both schools and homes. Usually the goal in schools is to keep the aircraft noise hourly equivalent sound level under 45 dB inside classrooms and the maximum sound level under 60 dB. The goal in homes is usually to attain a stated A-weighted noise reduction for aircraft noise. Typically, the stated amount of noise reduction is between 25 and 35 dB, varying with airport program and sometimes at various locations within a program. For the most part, the sound insulation is accomplished by replacing the windows with double glazing designed to provide good sound attenuation, improving ventilation as required, improving some doors, and sealing cracks and gaps with suitable caulking or gaskets. In many cases, sound insulation of homes is accompanied by an avigation easement that prevents the homeowner from suing the airport unless the noise increases significantly in the future.

Sound barriers around airports are usually earth berms with shrubbery and trees, free-standing walls, or buildings. For example, Minneapolis Airport has an extensive landscaped earth berm to reduce the noise from taxiing and idling aircraft at the near neighbor homes. It also improves the view toward the airport, translating an airport industrial complex scene into tranquil shrubs and trees. Walls are employed, particularly where distance between sound and receiver is small and land is intensively used. New York LaGuardia, Los Angeles International, and Boston Logan airports have sound barrier walls. Airport walls are usually much higher than highway barriers because of the height of the engines on the aircraft. At Logan, the walls are over 30 feet in height and a portion of the barrier consists of a multistory office building, sited for noise control purposes. Typical barriers are designed to attain a 10-dB or more reduction of the noise from ground operations at locations near to the barrier.

4.7 Change Land Use

Land use may be controlled through many means, including zoning restrictions, conversion to noise-compatible activities, purchase with compatible redevelopment, and purchase of avigation easements.[27]

New airports are typically built in areas where land is relatively cheap and for which the pre-airport land uses were low density and low value such as agricultural, surface mining, forestry, or otherwise not used. Once the airport is constructed, it creates a demand for commercial, industrial, and residential use. Most of the initial industrial use is often on the airport property and the commercial use is usually on and immediately adjacent to the airport property. Consequently, the highest use for the majority of the remaining surrounding land is residential development. Any attempt to restrain landowners from residential development is usually seen as depriving them of their rights. Therefore, zoning has proved to be ineffective in preventing residential development near airports.

Overlay zoning that imposes sound insulation requirements for new construction may be effective. However, purchase and redevelopment have proven most effective. In many cases, the land is incorporated into airport expansion or used for the development of industrial parks for commercial use. In others it may be used for recreational purposes such as public golf courses.

4.8 Monitor and Model

The implementation of a noise-monitoring program utilizing permanent and/or portable sound monitors with identification of aircraft and documentation of cumulative noise from aircraft and nonaircraft sources is essential to provide data for public discussion and validation of noise modeling. Where possible, this noise measurement program should be coordinated with a flight track monitoring program that provides aircraft identification and tracks to determine the causes of the various measured sounds. When a combined program is in place, designed with public participation and providing significant regular data output, it improves the possibility of improving public confidence in both the process and the results. These data then can be used as a basis for discussing the existing noise situation at an airport and examining alternatives that might provide future mitigation of noise.

4.9 Communication

All of the steps and progress toward the reduction of the physical measurable quantities of noise will not reduce the public's perception of the noise at an airport unless

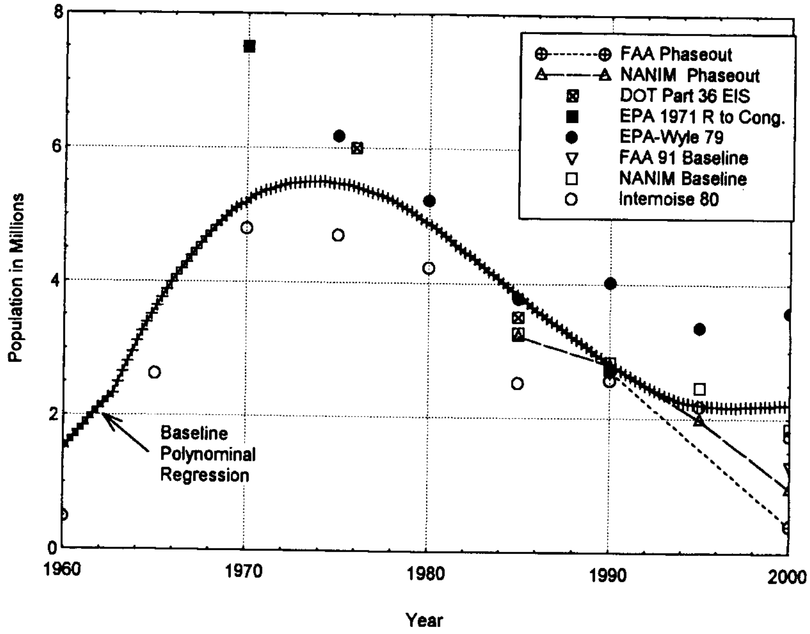


Fig. 5 Various estimates of the U.S. national population residing in neighborhoods where the L_{dn} exceeds 65 dB, illustrating the potential effect of the FAR Part 36 stage 2 phaseout in 2000 based on FAA[18] and NANIM[44] calculations and the estimates for the period 1960–2000 without the stage 2 phaseout from various sources (Refs. 42, 38, 43, 18, 44, and 40, respectively).

communication channels are established between the airport authorities, the local FAA Air Traffic Control personnel, the airport users, and the airport's neighbors. One of the most important outcomes of a successful Part 150 study is the development of a capability for the concerned parties to talk and to begin to understand each other. Open dialogue, consensus decision making, and teamwork in seeking creative solutions can build the beginnings of trust and respect among the parties and provide a basis for institutionalizing the process so that the parties continue to work together and develop a practical and rational decision-making capability.

5 CHANGES IN POPULATION IMPACT 1960–2000

The number of people residing in neighborhoods where the DNL exceeded 65 dB has been used as a measure of the total population impact of civil aircraft noise. This number rose from a small value in 1958 to a maximum in the early 1970s, then decreased as a result of the introduction of quieter aircraft and the implementation of airport-specific noise control measures. The maximum population impact is estimated to have been between 5 and 7.5 million people, occurring in the early 1970s.[38–42]

Figure 5 illustrates various estimates of population impact above a DNL of 65 dB.[38,40–43] The two phaseout curves are estimates of the effect of phasing out stage 2 aircraft in the year 2000, one made by the FAA in 1991[18] and the other by the National Airport Noise Impact Model[44] in 1987.[18,44] The baseline polynomial regression curve represents the points that do not assume a stage 2 phaseout in 2000. It shows a maximum of about 5.5 million people in 1973, a reduction to 2.2 million in 1995, with a slight increase thereafter. Both the FAA phaseout and National Airport Noise Impact Model (NANIM) phaseout curves coincide with the baseline polynomial regression in 1990 but decrease to 400,000–900,000 people in the year 2000. The FAA estimate[18] of 400,000 is probably the most accurate since it is based on the most recent data, particularly the results of the Part 150 studies at the nation's airports.

These estimates indicate that the extent of most intense noise impacts is in the process of being controlled. However, they do not address the number of people who have significant remaining noise impact at lower values of DNL, such as 60 dB, nor do they address the inevitable increase after the year 2000 after the noisiest aircraft are retired. Thus, unless the noise levels of new aircraft are reduced below the stage 3 noise limits, the growth of the fleet and its number of daily operations will ensure that the population suffering noise impact will increase.

REFERENCES

1. PNYA, "Air Terminals—Rules and Regulations II," Port of New York Authority adopted 1951 with Amendments, 1958.
2. "Noise Final Report," Wilson Report Cmnd.2056, London, 1966.
3. ANSI S12.9-1988, "American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound, Part 1," Acoustical Society of America, New York, 1988.
4. ANSI S1.1-1994, "American National Standard Acoustical Terminology," Acoustical Society of America, New York, 1994.
5. "Environmental Protection—Annex 16 to the Convention on International Civil Aviation, Vol. 1, Aircraft Noise," International Civil Aviation Organization, as amended, November 1985.
6. California, "Subchapter 6. Noise Standards," Register 79, No. 21-5-26-79, Public Utilities Code (Regulation of Airports), California Department of Transportation, Division of Aeronautics, with Amendments to March 1990.
7. S. Goldstein and A. H. Odell, "Comments on the Problem of Jet Aircraft Noise," Port of New York Authority, 1965.
8. J. P. Muldoon, "The Development of a New Aircraft Noise Abatement Monitoring System Using Air Traffic Control Radar Data," *Internoise 84 Proceedings*, Vol. 1, December 1984, pp. 689–692.
9. SAE AIR1845, "Procedure for Calculation of Airplane Noise in the Vicinity of Airports," Society of Automotive Engineers, A-21, March 1986.
10. FAA, "INM Integrated Noise Model Version 3 User's Guide," Federal Aviation Administration, FAA EE-81-17, October 1982.
11. C. L. Moulton, "Air Force Procedure for Predicting Aircraft Noise around Air Bases: Noise Exposure Model (Noise Map) User's Manual," U.S. Air Force Systems Comand, AAMRL-TR-90-011, February 1990.
12. Massport, "The Results of Four Years of Noise Abatement at Boston's Logan International Airport," Massport Noise Abatement Office, 1981.
13. Federal Aviation Regulations Part 36, "Noise Standards: Aircraft Type and Airworthiness Certification," Federal Aviation Administration, November 1969.
14. Federal Aviation Regulations Part 36, "Noise Standards: Aircraft Type and Airworthiness Certification," Federal Aviation Administration, Amendment 7, February 1977.
15. FAA, Advisory Circular AC36-1F, "Noise Levels for U.S. Certificated and Foreign Aircraft," Federal Aviation Administration, June 1992.
16. K. M. Eldred, "Airport Noise: Solving a World Class Problem," *Proceedings, Internoise 92*, Vol. 1, July 1992, pp. 3–12.
17. U.S. Congress, "Airport Noise and Capacity Act of 1990," PL101-508, November 1990.
18. FAA, "Transition to an All Stage 3 Fleet Operating in the 48 Contiguous United States and the District of Columbia," Federal Aviation Administration, 14 CFR Part 91 Amendment, September 1991.
19. U.S. Congress, "National Environmental Protection Act of 1969," PL91-190, 42 U.S.C. 24321, 1969.
20. FAA Order 1050.1D, "Policies and Procedures for Considering Environmental Impacts," Federal Aviation Administration, December 1983.
21. U.S. Congress, "Noise Control Act of 1972," PL92-574, 42 U.S.C. 4901, with amendments, 1977.
22. U.S. Congress, "Aviation Safety and Noise Abatement Act of 1979," PL101-508, 1990.
23. FAA Order 5050.4A, "Airport Environmental Handbook," Federal Aviation Administration, October 1985.
24. FICUN, "Guidelines for Considering Noise in Land Use Planning and Control," Federal Interagency Committee on Urban Noise," June 1980.
25. FICON, "Federal Agency Review of Selected Airport Noise Analysis Issues," Federal Interagency Committee on Noise, Environmental Protection Agency, Washington, D.C., 1992.
26. FAA, "Airport Noise Compatibility Planning," 14 CFR Part 150, as amended, March 1988.
27. FAA Advisory Circular 150/5020-1, "Noise Control and Compatibility Planning for Airports," Federal Aviation Administration, August 1983.
28. FAA, "Report to Congress Part 150 Airport Noise Compatibility Planning," Federal Aviation Administration, November 1989.
29. FAA, "Notice and Approval of Airport Noise and Access Restrictions," FAR Part 161, Federal Aviation Administration, September 1991.
30. FAA, ANE-500-79-3, "Final EIS Departure Procedures, Runway 22 Right, Logan International Airport," March 1980.
31. FAA Order 8400.9, "National Safety and Operational Criteria for Runway Use Programs," November 1981.
32. K. M. Eldred, "Massport Phase II Preferential Runway Study for Logan International Airport," Ken Eldred Engineering Report 82-80, Concord, MA, November 1980.
33. FAA Advisory Circular 36-3F, "Estimated Airplane Noise Levels in A-Weighted Decibels," Federal Aviation Administration, August 1990.
34. FAA Advisory Circular 36-2C, "Measured or Estimated (Uncertificated) Airplane Noise Levels," Federal Aviation Administration, February 1986.
35. ICAO Circular 157-AN/101, "Assessment of Technological Progress Made in Reduction of Noise from Subsonic and Supersonic Jet Aeroplanes," Montreal, 1981.
36. "The Noise Benefits Associated with the Use of Continuous Descent Approach and Low Power/Low Drag Approach Procedures at Heathrow Airport, CAA Paper 78006.
37. FAA Advisory Circular AC91-53.A, "Noise Abate-

ment Departure Profile," Federal Aviation Administration, 1994.

38. EPA, "Report to the President and Congress on Noise," Senate Document No. 92-63, February 1972.
39. DOT-NASA, "Civil Aviation Research and Development Policy Study," Joint DOT-NASA Report DOT TST 10-5, NASA SP266, March 1971.
40. K. M. Eldred, "Model for Airport Noise Exposure on a National Basis—1960 to 2000," *Proceedings Internoise 80*, December 1980, pp. 803–808.
41. J. E. Wesler, "Airport Noise Abatement—How Effective Can It Be?" *J. Sound Vib.*, February 1975, pp. 16–21,
42. DOT, FAA, "FAR Part 36 Compliance Regulation, Final EIS," Federal Aviation Administration, November 1976.
43. C. Bartel and L. C. Sutherland, "Noise Exposure of Civil Air Carrier Airplanes Through the Year 2000," EPA 550/9-79-313-1, February 1979.
44. K. M. Eldred, "The National Airport Noise Impact Model and Its Application to Regulatory Alternatives," FAA EE88-3, July 1987.

88

SURFACE TRANSPORTATION NOISE

ROBERT HICKLING

1 INTRODUCTION

Surface transportation noise is caused by highway vehicles such as trucks, buses, passenger cars, and motorcycles, together with railroad vehicles, both above and below ground. Also included is the noise of off-highway vehicles, used for construction, agri- and horticulture, and recreation. The noise of boats, ships, and hover craft can also be regarded as surface transportation noise. Although the noise of wheeled vehicles in cities has been of concern for centuries,[1] it is only in recent decades that quantitative noise measurements have been developed. Major government reports in the 1960s and 1970s[2,3] indicated the need for noise standards. The principal burden of meeting the vehicle noise standards that eventually developed fell on vehicle manufacturers. Highway vehicle manufacturers with worldwide markets are required to meet a diversity of exterior or community noise standards. Interior noise standards, developed as a result of studies of hearing loss due to noise exposure in the workplace (see Chapters 86 and 119), are by their nature less diverse. Since noise in the workplace is covered in Chapter 86, the emphasis here is on exterior noise.

2 NOISE SOURCES IN SURFACE TRANSPORTATION

The same noise sources occur in different forms of surface transportation. These are shown, for example, in Fig. 1 for a heavy-duty truck and a motorcycle. Primarily there is the noise associated with the power source, which usually is an air-breathing diesel or gasoline engine. This consists of noise from the air intake, exhaust, cooling system, engine structural vibrations, pumps for water, fuel and lubricating oil, and electrical and other ancillary systems. Diesel (compression ignition) engines are noisier than gasoline (spark ignition) engines because compression ignition involves a greater impulsive force. Diesel engines are built more strongly than gasoline engines to withstand this force. Diesels, especially direct-injection diesels, are more efficient and are used for heavy-duty trucks and diesel–electric locomotives, where they power the electric motors that drive the wheels on the rails. Gasoline engines have a wider speed range than diesels, requiring fewer gears to operate them. Diesels in passenger cars generally have indirect injection into a precombustion chamber, which reduces noise and provides a speed range comparable to gasoline engines. Indirect-injection diesels are more efficient than gasoline engines but less efficient than direct-injection diesels. Gasoline engines are preferred in passenger cars because they are less noisy and lack the characteristic diesel odor.

Noise from the drive train comes from the transmission, drive axle, differential gears, and wheel bearings. There is also noise associated with surface propulsion, that is, the interaction of tires with the road, and of wheels with rails, track systems with the ground, and marine propellers and hulls in water. Noise originates also from electrical generators and alternators and from air conditioning, power steering, and braking systems. Brake squeal continues to be a chronic problem.[4,5] In addition, there is noise due to airflow over the vehicle. This occurs principally in interior noise; however, it can be a significant part of exterior noise, especially at higher speeds.

To develop methods of controlling the noise of surface transportation, it is necessary to understand how

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker
Copyright © 1997 John Wiley & Sons, Inc.

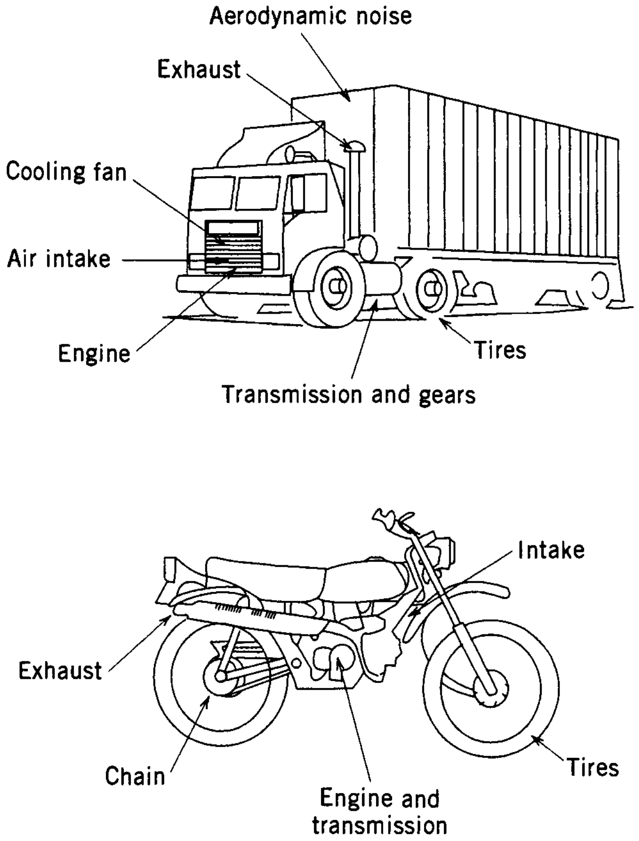


Fig. 1 Noise sources in a heavy-duty truck and a motorcycle. (After Ref. 1.)

the different vehicle components generate noise and how they are designed. The total sound power emanating from a vehicle is miniscule compared to the power involved in the operation of the vehicle. For example, the sound power of a passenger car at constant speed under road load is several milliwatts, whereas the power involved in operating the vehicle is of the order of kilowatts. Noise reduction is therefore not always of prime importance in the design of surface transportation, which is unfortunate because the understanding required to reduce noise can result in other improvements. For example, the design of a quieter passenger car cooling fan resulted in improved fan efficiency[6] with no significant increase in cost. Control of engine and machinery noise is discussed in a number of texts.[7–9] The development of sound intensity techniques within the past two decades (see Chapter 156) is important for vehicle noise diagnostics and standard tests. Also active control (see Chapter 75) provides a new approach to reducing noise and vibration, for example, reducing vehicle interior noise.[10]

With new lower limits in vehicle passby noise standards, tire noise has become an increasingly important contributor to the overall noise of a vehicle, comparable to intake and exhaust noise. The noise due to tire–road interaction has received much attention.[11,12] Regulation of tire noise separate from the rest of the vehicle has been proposed.[13] The relation between tire noise and the exterior aerodynamic noise of a vehicle has been discussed.[14] Aerodynamic noise is roughly proportional to the sixth power of air speed. Hence, as vehicle speed increases, aerodynamic noise eventually predominates over tire noise. Figure 2 shows how the aerodynamic noise of a vehicle can be separated from tire noise. In the figure, the tire noise of a coasting stake-bed truck is kept constant by keeping the ground speed at 80 km/h, while the air speed is varied by coasting the vehicle with and against a prevailing wind. Sound intensity measurement close to a tire on a roadway has provided useful information on how tire noise is generated.[15] This type of measurement could also be used as a standard tire noise test to replace coast-by tests.[15,16] Standardization of test track surfaces is being proposed.[17] Tire noise has been compared under various conditions of cruise and acceleration.[18]

Wheel–rail interaction is an important contributor to railroad noise,[19] caused mainly by the weight of rolling stock and metal-to-metal contact. Roughness of the wheel and rail play[5] an important role. It is believed that most of the noise is radiated by the rails.[20] A significant part of the noise can be caused by vibrations transmitted to rolling-stock structures because the suspension system generally goes not provide good isolation from wheel–rail interaction. Structures connected to the rails also generate noise. Elevated railroads are notoriously noisy, as are trains passing over bridges. In addition there is rumble in buildings near a railroad caused by the transmission of low-frequency vibrations through the ground.[21]

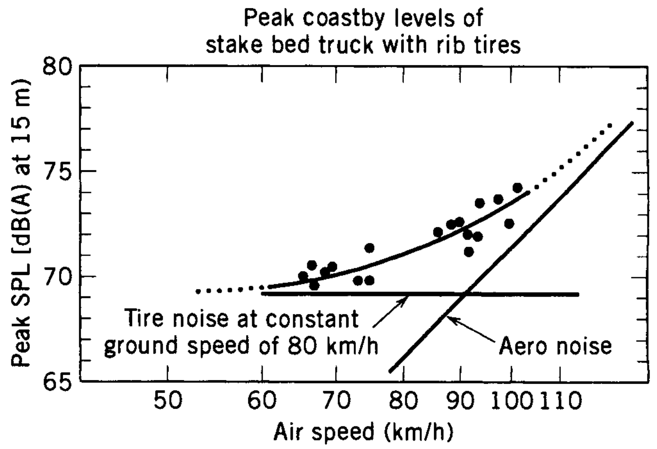


Fig. 2 Separation of aerodynamic noise from tire noise of a coasting vehicle using the prevailing-wind method. (From Ref. 2.)

3 TEST PROCEDURES AND LIMITS FOR EXTERIOR NOISE

Noise regulations are outlined for three main economic regions, the United States and Canada, the European Union, and Japan. Vehicle noise regulations in other countries are also briefly discussed. Vehicle noise regulations overseas are much more stringent than in the United States.

3.1 Noise Regulations in the United States and Canada

Table 1 lists the noise limits that apply to new highway vehicles sold in the United States and Canada. The U.S. government regulates the noise of medium- and heavy-duty trucks, including motor homes, and also motorcycles and mopeds. Passenger cars, buses, and other light vehicles are regulated by certain states and local jurisdictions whose combined effect determines the noise level of new light vehicles sold in the United States. Light vehicles are generally defined as having a gross vehicle weight rating (GVWR) less than 4.5 tonnes. The GVWR is the loaded weight of a vehicle, as specified by the manufacturer. The passby test procedure for light vehicles in the United States and Canada is the SAE J986.[22] For heavy vehicles over 4.5 t, the test is the SAE J366, equivalent to Section 40, Part 205, Subpart B, of the Code of Federal Regulations. For motorcycles and mopeds, the test is the SAE F76a, equivalent to Section 40 CFR, Part 205, Subpart D, of the Code of Federal Regulations; Subpart E presents noise-labeling requirements for motorcycle replacement exhaust silencers. All these tests are maximum-acceleration tests past a microphone 15 m from the center of the lane of travel of the vehicle and 1.2 m above the ground.

The passby site for the SAE J986 test is shown in Fig. 3. At the start point, maximum acceleration is applied, and the approach to the start point is chosen so that the vehicle can reach maximum rated engine speed in the end zone. Maximum noise during the passby is recorded. Originally, the test was run in first (lowest) gear; however, test constraints now force many vehicles to be tested in second gear. Ambient noise has to be at least 10 dB(A) below the passby reading, and wind speeds have to be less than 19 km/h. The test has to be run on a dry roadway, without dirt or powdery snow. Both sides of the vehicle are tested. Four "good" runs are required, that is, readings within 2 dB(A) of each other for each side of the vehicle, which are averaged arithmetically. The average of the two highest readings for the noisiest side of the vehicle is the test result.

The regulations for railroad noise are contained in Section 40, Part 201, and Section 49, Part 210, of the Code of Federal Regulations. Test measurements are made within a clear level area 30 m from the track center line. Locomotives are tested either stationary using remote load cells or in motion. Noise limits for locomotives manufactured in 1980 and thereafter are as follows:

Stationary, idle throttle setting	70 dB(A)
Stationary, all other throttle settings	87 dB(A)
Moving	90 dB(A)

For rail cars, the limits are as follows:

Moving at speeds 83 km/h or less	88 dB(A)
Moving at speeds greater than 83 km/h	93 dB(A)

Weather and ambient noise restrictions apply as stated above for the J986 test.

3.2 Noise Regulations in the European Union

Table 2 lists the exterior limits that apply to new highway vehicles sold in the European Union (EU), or, as it was previously known, the European Economic Community (EEC). These limits are mandatory and should not be confused with the regulations of the United Nations Economic Commission for Europe (UN/ECE, or ECE), which are voluntary on the part of participating countries. It is planned to harmonize EU and ECE noise regulations in the future. The EU vehicle noise test procedure is provided in Council Directives 92/97/EEC and 81/334/EEC, as presented in the Official Journal of the European Communities. For automatic-drive and high-power vehicles, this was amended to 84/372/EEC. The test is based on ISO 362, which is equivalent to SAEJ1470.[22] A 1-dB(A) tolerance is permitted in the test. This test and its derivatives are widely used throughout the world. A good discussion of the test is presented in Ref. 23. The test site configuration is shown in Fig. 4. It is used for all classes of highway vehicles, including motorcycles. It is important to note the difference in the size of the test site, compared to the larger J986 test site in Fig. 3. In particular, the measurement distance is 7.5 m compared to 15 m in the J986 test. The J986 test is designed to measure the maximum noise output of a vehicle, whereas the EU test is designed to measure the noise of a typical urban acceleration. As the limits in Table 2 are reduced, however, the two measurement objectives approach each other. The fact that the measurement distance is half that of the J986 test is considered to imply that, for the same vehicle, the EU test gives results that are about 6 dB(A) higher than the results of the J986 test. However, the difference between the tests is more complicated than this, and in practice the correlation is sometimes quite poor.

TABLE 1 Summary of Highway Vehicle Noise Limits in the United States and Canada

Jurisdiction	PC	MPV	LT	HT	BUS	SBUS	MH	MC	MOP	GVRW (T)	Static dB(A)	Passby dB(A)
U.S. federal									x			70
								x				80
				x			x			> 4.5		80
Canada[a]								x				80
	x	x	x							<= 2.7		80
		x	x	x	x	x	x			<= 4.5		83
				x	x	x	x			> 4.5		83
California	x	x	x							<= 4.5	95	80
				x	x	x	x			> 4.5		80
Colorado	x	x	x							< 2.7		84
		x	x	x	x	x	x			>= 2.7		86
Florida[b]	x	x	x							<= 4.5	95	80
		x								<= 4.5	95	83
				x	x	x	x			> 4.5		83
Maryland[b]	x	x	x							<= 4.5		80
				x	x	x	x			> 4.5		83
Nebraska[c]				x						>= 4.5		80
					x	x				> 4.5		83
Nevada	x	x	x							< 2.7		84
		x	x	x	x	x	x			>= 2.7		86
Oregon[b]	x	x	x							<= 4.5	93/95[d]	80
				x	x	x	x			> 4.5		80
Washington	x	x	x							<= 4.5		80
				x	x	x	x			> 4.5		80
Cook County, IL[b]	x	x	x							<= 4.5		80
				x	x	x	x			> 4.5		83
Boston, MA	x	x	x							<= 4.5		80
				x	x	x	x			> 4.5		80
Chicago, IL (w/in Cook County)	x	x	x							<= 4.5		80
				x	x	x	x			> 4.5		80
Des Plaines, IL (w/in Cook County)[b]	x	x	x							<= 4.5		80
				x	x	x	x			> 4.5		84
Grand Rapids, MI	x	x	x							<= 4.5		80
				x	x	x	x			> 4.5		84
Madison, WI	x	x	x							<= 2.7		84
		x	x	x	x	x	x			> 2.7		86
Prairie Village, KS	x	x	x							< 3.6		75
		x	x	x	x	x	x			>= 3.6		75
Urbana, IL	x	x	x							< 3.6		80
		x	x	x	x	x	x			>= 3.6		74
Washington, DC	x	x	x							<= 4.5		80
				x	x	x	x			> 4.5		80

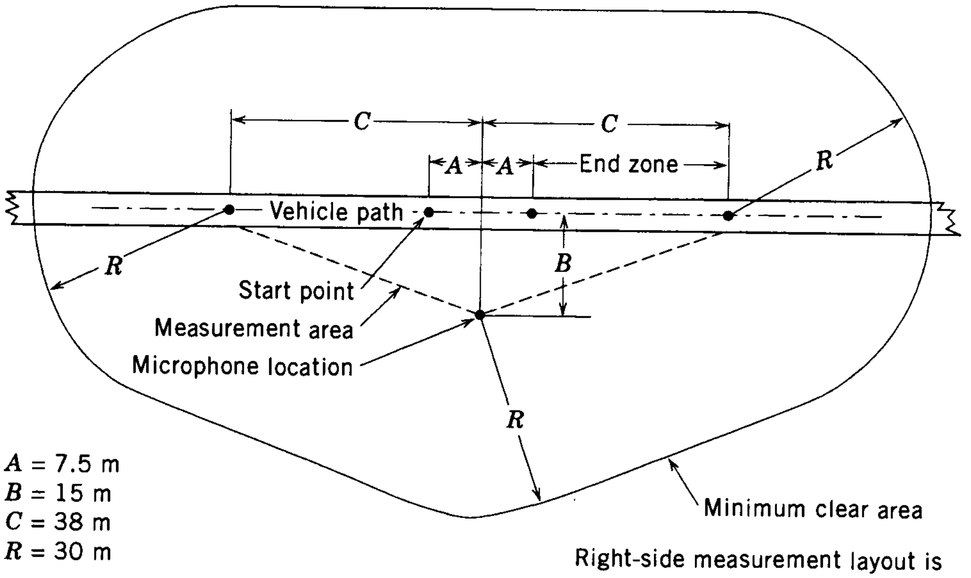
Abbreviations: PC, passenger car; MPV, multipurpose vehicle; LT, light truck; HT, heavy truck; MOP, moped; BUS, transit/intercity bus; SBUS, school bus; MH, motor home; MC, motorcycle.

[a] Canadian noise regulations allow a 2-dB (A) tolerance on noise limits.

[b] Certification required

[c] Applies to diesel engine vehicles only.

[d] 2500 r/min; 93 dB(A) and front-engine vehicles; 95 dB(A), rear and midengine vehicles.



Right-side measurement layout is shown. Start point and end zone are interchanged for left-side measurement.

Fig. 3 Plan view of test site for the SAE J986 test. (From SAE Standard J986. Reproduced with permission from the Society of Automotive Engineers, Warrendale, PA.)

At present the EU has no exterior noise standards for railroad equipment because of intermingling with equipment from outside the EU and differences in gauge size. Railroads use a number of noise standards in an ad hoc manner, such as the ORE E82/RP.4 developed by the International Union of Railways and the ISO 3095: 1975.

An important development in EU noise standards is the use of sound power as a noise measure, instead of sound pressure at a single point in space, for construction site and earth-moving equipment. Sound power measures the total noise emanating in all directions from a source. Details are given in directives 95/27/EU, 89/514/EEC, 86/662/EEC, 84/533-536/EEC, and 79/113/EEC. From 79/113/EEC, the prescribed distribution of measurement positions on a hypothetical measurement hemisphere enclosing a stationary piece of earth-moving machinery is shown in Fig. 5. Sound pressure is measured at the measurement points, which is converted to sound intensity, using the far-field approximation and integrated to obtain sound power. The sound power determination would be more accurate[24–26] if sound intensity measurements were used (see Chapter 156) instead of sound pressure. In its present form, the test uses only half the measurement points shown in Fig. 5. A similar sound power test was developed for lawn mowers in 84/538/EEC. New noise limits in the EU for earth-moving machines of

TABLE 2 European Union Highway Vehicle Noise Limits

Vehicle Description	Unloaded Weight (t)	Power (kW)	Limit [dB(A)]
Passenger car	—	—	74
Mini bus >9 seats	<2	—	76
	>2 < 3.5	—	77
Bus >9 seats	>3.5	<150	78
Light truck/van	<2	—	76
	>2 < 3.5	—	77
Medium truck/van	>3.5	<75	78
	>3.5	>75 < 150	78
Heavy trucks	>12	>150	80
Motorcycles			
≤80 cc			75
>80 ≤ 175 cc			77
>175 cc			80

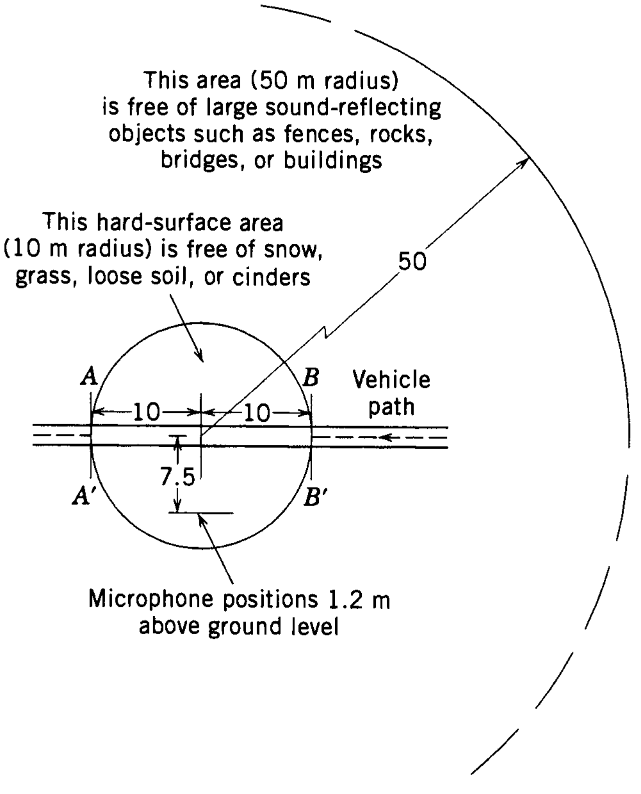

This area (50 m radius) is free of large sound-reflecting objects such as fences, rocks, bridges, or buildings
This hard-surface area (10 m radius) is free of snow, grass, loose soil, or cinders
50
A
B
Vehicle path
10
10
7.5
A′
B′
Microphone positions 1.2 m above ground level

Fig. 4 Plan view of test site for the ISO 362 test. All dimensions in metres. (From SAE Standard J1470. Reproduced with permission from the Society of Automotive Engineers, Warrendale, PA.)

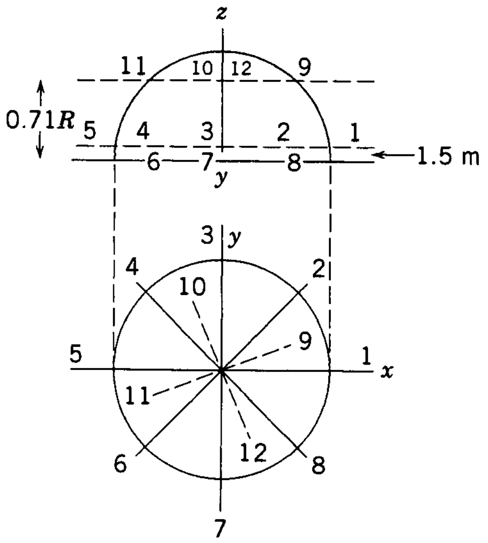

z
11
10
12
9
0.71R
5
4
3
2
1
6
7
8
y
1.5 m
3
y
4
10
2
5
9
1
x
11
6
12
8
7

Fig. 5 Distribution of measurement points on hypothetical measurement hemisphere for determining the sound power of earth-moving equipment, as prescribed in 79/13/EEC.

net installed power less than 500 kW come into effect at the end of 1996. Machines with power greater than 500 kW operate in quarries and mines and are considered to have a negligible effect on community noise. Between 1996 and 2001, permissible sound power levels L_{WA}, in A-weighted decibles relative to 1 pW, are given by the following formulas:

1. Tracked vehicles (except excavators):

$$L_{WA} = 87 + 11 \log P \quad \text{for } L_{WA} \geq 107.$$

2. Wheeled dozers, loaders, excavator-loaders:

$$L_{WA} = 85 + 11 \log P \quad \text{for } L_{WA} \geq 104.$$

3. Excavators:

$$L_{WA} = 83 + 11 \log P \quad \text{for } L_{WA} \geq 96.$$

In the preceding P is the net installed power of the vehicle in kilowatts. Below the lower limit on the right of the formulas the machine automatically passes the test. After 2001 all the numerical A-weighted decibel values in the above formulas, including lower limits, are reduced by 3 dB(A). The coefficient 11 is unchanged.

3.3 Vehicle Noise Regulations in Japan

In Japan, exterior noise tests for road vehicles were used as far back as 1971.[27] The present tests are very thorough and consist of three parts[28]:

1. A fixed-speed passby test at 7 m, 60% rated engine speed or 65 km/h
2. An acceleration passby test at 7.5 m
3. A stationary test behind the exhaust outlet at 20 m

In all the tests the microphone is 1.2 m above ground level. The acceleration passby test is essentially the same as the EEC test; however, there is no 1-dB(A) tolerance and the test applies only to the left side of the vehicle. Noise limits for the acceleration test are given in Table 3. The limits for passenger cars, medium- and heavy-duty trucks, and buses will be reduced by 2 dB(A) in 1998. The stationary test limit is 103 dB(C). Japan also has a close-in stationary exhaust noise test based on ISO 5130. Reductions in these limits are planned for 1998 and 2002.

TABLE 3 Noise Limits for New Highway Vehicles in Japan

Vehicle Description	GVWR (t)	Power (kW)	Limit [dB(A)]
Passenger car	—	—	78
Light truck	<3.5	—	81
Medium truck, bus	>3.5	<150	83
Heavy bus	>3.5	>150	83
Heavy truck	>3.5	>150	83
Moped < 50 cc			72
Motorcycle			
>50 cc <125 cc			72
>125 cc <250 cc			75
>250 cc			78

TABLE 4 Sample of Passenger Car Noise Limits in Different Countries

Country	Limit[a] [dB(A)]
Austria	74
Switzerland	74
Finland	74
Norway	74
Sweden	74
Australia	77
Czechoslovakia	74
Hungary	80
Israel	80
Korea	75
Gulf States	82

[a] No 1-dB(A) tolerance.

3.4 Vehicle Noise Regulations in Other Countries

Many other countries have vehicle noise regulations, but it is not possible to give these here in detail. The information can be obtained from appropriate government agencies. A sampling of passenger car exterior noise limits is presented in Table 4. These are based on the EEC/ECE test procedures at 7.5 m. As the allowable limit drops, tire noise makes a proportionately increased contribution to the passby noise level.

4 VEHICLE INTERIOR NOISE

As previously stated, regulation of vehicle interior noise is based on regulation of noise in the workplace (see Chapter 86) and is applied principally to noise inside truck and locomotive cabs. In the United States, locomotive cab noise is regulated by the Federal Railroad Administration, as specified in 49 CFR 229.121, with allowed levels of noise exposure as given in Table 5. Tests are conducted 150 mm from a crew member's ear using the procedures described in Chapter 86. Truck cab noise is regulated by the Federal Highway Administration, as specified in 49 CFR 393.94. The test is conducted with the engine operated in neutral at maximum rated speed. The sound level limit is 90 dB(A), measured 150 mm to the right of the driver's ear. By current standards (see Chapter 119), these levels are too high for adequate protection of hearing. In passenger vehicles, both on highways and on railroads, control of noise is governed principally by customer complaint and commercial competition. Recently concern has developed over exterior noise from auto stereo systems.

The sources in a vehicle that excite interior noise are the same as those discussed in Section 2. Generally a distinction is made between airborne or structure-borne excitation. The fact that interior noise occurs within a closed structural shell makes a significant difference. In particular, interior noise is aggravated when an external excitation couples with a cavity resonance of the structure. This occurs principally at low frequencies. Airflow over the vehicle is a significant contributor to interior

TABLE 5 Duration Limits for Locomotive Crew

Duration (h)	Sound Level [dB(A)]
12	87
8	90
6	92
4	95
2	100
1.5	102
0.5	110
0.25 or less	115

noise, especially in light vehicles. Wind tunnel tests to reduce the aerodynamic drag of light-vehicle prototypes are usually combined with tests to measure aerodynamic interior noise.

5 CONCLUDING COMMENTS

Surface transportation is the major contributor to community noise. In coming to grips with the problem, a diversity of standard test procedures has developed throughout the world. Vehicle manufacturers would prefer to test indoors to avoid delays and additional cost due to bad weather. Large expensive anechoic facilities have been built to simulate the passby test indoors. The test vehicle is operated on a dynamometer and the passby acceleration is simulated by recording sequentially along a row of microphones. This simulation has been accepted by regulatory agencies as a substitute for the outdoor passby test. Passby tests determine a somewhat ill-defined quantity that depends on the acoustic radiation pattern of the vehicle and on the type of test. Obviously it would be better to standardize on a common, well-defined quantity, such as sound power. Indoor sound power tests of passenger cars and a truck have been performed in the United States.[24–26] Such tests provide data with much less variability than the outdoor passby test. Also they provide diagnostic information that can be used to identify and correct noise problems. In addition, the sound power of the tires on the dynamometer can be subtracted from the total sound power of the vehicle. Indoor sound power tests are performed routinely in the United States on earth-moving equipment, to meet European noise standards.[29]

REFERENCES

1. M. J. Crocker, "Noise of Surface Transportation to Nontravelers," in D. N. May (Ed.), *Handbook of Noise Assessment*, Van Nostrand Reinhold, New York, 1978.
2. Wilson Committee, *Noise—Final Report*, H. M. Stationery Office, London, 1963.
3. Administrator of the Environmental Protection Agency, *Report to the President and Congress on Noise*, U.S. Government Printing Office, Washington, D.C., 1972.
4. R. A. C. Fosbery and Z. Holubecki, "Disc-Brake Squeal: Its Mechanisms and Suppression," Research Rept. No 1961/2, Motors Industries Research Association (MIRA), Nuneaton, Warks, UK, 1961.
5. G. D. Liles, "Analysis of Disc-Brake Squeal Using Finite-Element Methods," SAE Paper 891150, Society of Automotive Engineers, Warrendale, PA, 1989.
6. R. E. Longhouse, "Control of Tip-Vortex Noise of Axial-Flow Fans Using Rotating Shrouds," *J. Sound Vib.*, Vol. 58, No. 1, 1978, pp. 201–214.
7. R. Hickling and M. M. Kamal (Eds.), *Engine Noise: Excitation, Vibration and Radiation*, Plenum, New York–London, 1982.
8. D. Baxa (Ed.), *Noise Control in Internal Combustion Engines*, Wiley, New York, 1982.
9. R. H. Lyon, *Machinery Noise and Diagnostics*, Butterworths, Boston, 1987.
10. P. Rogers and C. R. Fuller (Eds.), *Recent Advances in Active Control of Sound and Vibration*, Technomic, Lancaster, PA, 1991.
11. R. Hillquist (Ed.), "Highway Tire Noise," *Proceedings of SAE Symposium*, San Francisco, CA, Society of Automotive Engineers, Warrendale, PA, 1976.
12. K. Samberg, "Tire/Road Noise," *Proceeding of 2nd International Tire/Road Noise Conference*, Gothenburg, Sweden, Vols. I and II, STU/Information Nos. 794, 795/1990, Swedish National Board for Technical Development, Stockholm, 1990.
13. W. H. Close, "Should There Be Truck Tire Noise Regulations?" *Sound and Vibration*, Feb. 1975.
14. R. Hickling and L. J. Oswald, "A Proposed Revision of the SAE J57a Tire Noise Testing Procedure," General Motors Research Laboratories, Warren, MI, Research Publication No. GMR 2816, 1978.
15. P. Donavan and L. J. Oswald, "Identification and Quantification of Truck Tire-Noise Sources under Road-Operating Conditions," *Proceedings INTER-NOISE 80*, Miami, FL, 1980.
16. J. S. Bolton and H. H. Hall, "Correlation of Tire Intensity Levels and Passby Sound Pressure Levels," SAE Paper No 951355, Society of Automotive Engineers, Warrendale, PA, 1995.
17. U. Sandberg, "Standardization of a Test Track Surface for Use during Vehicle Noise Testing," SAE Paper No. 911048, Society of Automotive Engineers, Warrendale, PA, 1991.
18. P. R. Donavan, "Tire-Pavement Interaction Noise Measurement Under Vehicle Operating Conditions of Cruise and Acceleration," SAE Paper 931276, Society of Automotive Engineers, Warrendale, PA, 1993.
19. C. Stanworth, "Sources of Railway Noise," in P. M. Nelson (Ed.), *Transportation Noise Reference Book*, Butterworths, Boston, 1987, Chap. 14.
20. P. J. Remington, "Wheel/Rail Noise: The State of the Art," *NOISE CON 77*, Hampton, VA, 1977, pp. 257–284.
21. P. J. Remington, L. G. Kurzweil, and D. A. Towers, "Low-Frequency Noise and Vibration from Trains," in P. M. Nelson (Ed.), *Transportation Noise Reference Book*, Butterworths, Boston, 1987, Chap. 16.
22. *Handbook of the Society of Automotive Engineers*, Warrendale, PA, 1991.
23. D. Morrison, "Road Vehicle Noise-Emission Legislation,"

in P. M. Nelson (Ed.), *Transportation Noise Reference Book*, Butterworths, Boston, 1987, Chap. 9.

24. R. Hickling, "Narrow-band Indoor Measurement of the Sound Power of a Complex Mechanical Noise Source," *J. Acoust. Soc. Am.*, Vol. 87, 1990, pp. 1182–1191.
25. J. Pope, R. Hickling, D. A. Feldmaier, and D. A. Blaser, "The Use of Sound-Intensity Scans for Sound-Power Measurement and for Noise-Source Identification in Surface Transportation Vehicles," SAE Paper No. 810401, Society of Automotive Engineers, Warrendale, PA, 1981.
26. R. Hickling, P. Lee, and W. Wei, "Investigation of Integration Accuracy of Sound-Power Measurement Using an Automated Sound-Intensity System," *Appl. Acoust.*, in press.
27. Automobile Type Approval System in Japan, Sept. Traffic Safety and Nuisance Research Institute, Ministry of Transport, Japan, TRIAS 20, 1971, 1977, p. 579.
28. Japan Ministry of Transport, Automobile Type Approval Test Division, Partial Revision of Noise Test Procedure for Road Vehicles, TRIAS 20, 1980.
29. L. Tweed, "Implementation of a Second-Generation Sound-Power Test for Production Testing of Earthmoving Equipment," *Proceedings of SAE Noise and Vibration Conference*, Society of Automotive Engineers, SAE Publication P222, 1989, pp. 185–192.

89

COMMUNITY RESPONSE TO ENVIRONMENTAL NOISE

SANFORD FIDELL AND KARL S. PEARSONS

1 INTRODUCTION

Transportation (air, road, and rail) noise is the dominant and most common source of noise exposure* in residential neighborhoods, and annoyance is its most consequential effect. Since noise is generally defined as "unwanted sound," it follows almost as a matter of definition that annoyance is among the most immediate and obvious effects of noise exposure on people: who, after all, enjoys "unwanted" exposure to any agent? No other effect of noise on communities (e.g., speech, sleep, and task interference) is as widespread, long-lasting, and well documented as annoyance. Residential exposure to community noise sources poses no meaningful risk of hearing damage nor any credible risk of extraauditory health effects.

Annoyance is not an immediate sensation such as loudness, governed solely by characteristics of acoustic signals, nor is it an overt behavior such as a complaint. Annoyance is instead an attitude with both acoustic and nonacoustic determinants. If noise is defined less elliptically as "acoustic energy that someone else considers too inconvenient or too expensive to control," it is apparent that an attitude engendered by noise exposure must have both acoustic and nonacoustic determinants. A brief review of community noise research provides background useful for understanding modern approaches to predicting the prevalence of annoyance in communities.

2 BACKGROUND

It has long been understood that the acoustic determinants of annoyance can include all of the factors that affect the audibility of sounds by people. The first community noise survey[1] provided an early demonstration of this principle by establishing the need for frequency weighting networks to account for human response to differing distributions of acoustic energy in the frequency domain. This pioneering work showed that measurement of noise in a manner relevant to its effect on people cannot be arbitrarily based on the bandwidth, frequency response, and dynamic range of available instrumentation. The noise worth measuring is the noise that people can hear. Creation and standardization of the A, B, and C weighting networks in the 1930s by inversion of the Fletcher–Munson contours[2] was the first embodiment of this principle.

Interest in systematic means for predicting community response to aircraft noise can be traced at least to the early 1950s. Rosenblith and Stevens[3] developed a Com-

---

*The term *exposure* is commonly used in two ways. One use of the term implies the time integral of intensity, while the other use implies the average sound intensity over a specific time period. Intensity is the rate of flow of sound energy per unit area per second. At distances from sound sources that are of interest in environmental analyses, sound intensity is directly proportional to the square of sound pressure. Thus, sound exposure is usually represented as the time integral of squared sound pressure. This process is often referred to informally as *energy summation*. Magnitudes are reported in logarithmic terms. For example, sound exposure level is 10 times the logarithm to the base 10 of the ratio of sound exposure to a reference exposure of 400 $\mu Pa^2 \cdot s$. In this logarithmic form, squared sound pressure is called sound level and expressed in units of decibels. Sound level in decibel notation is often expressed as an average (equivalent) sound level over a specified time interval (usually 1 h or 24 h). Single events are often described by their sound exposure level (SEL) with a reference time interval of one second.

Encyclopedia of Acoustics, Volume Two. Malcolm J. Crocker

Copyright © 1997 John Wiley & Sons, Inc.

munity Noise Rating (CNR) based on two tacit assumptions: (1) that the prevalence of annoyance grows monotonically with the loudness of noise exposure and (2) that both acoustic and nonacoustic factors affect annoyance. CNR attempted to account for community reaction via "level ranks," an estimate of the effective loudness of noise exposure. These level ranks were modified by more or less arbitrary adjustments of 5 dB for nonacoustic factors such as novelty and economic dependence on noise sources. A nonquantitative role was even provided for ambient noise levels, on the assumption that masking of low-level sounds could reduce their audibility and hence annoyance. The pragmatic approach of this study set the style for investigations in this field for many years.

Concern about noise pollution in residential areas intensified greatly when jet transport aircraft began commercial operation in the late 1950s. Questions were raised in airport neighborhoods about the compatibility of airports and communities, about limits on the growth of the air transport industry, about measures that could be taken to mitigate airport noise impacts, and about tolerable levels of noise exposure. Within a few years of the start of jet air transport service, an initial round of large-scale social surveys had been conducted in the vicinity of major international airports. (See Schultz[4] for a comprehensive discussion of early survey results and noise measurement methods.)

Little standardization of community response survey methods, questionnaire items, noise measurements, or analysis techniques was practiced throughout the 1960s and 1970s. Nonetheless, interpretations made in the early 1960s of the findings of the first generation of social surveys alarmed many. Aircraft noise exposure was viewed by some as seriously impairing important aspects of residential living. At the same time, many laboratory studies were undertaken of the annoyance, speech, and sleep interfering properties of noise. There was essentially no agreement on which acoustic and which social variables were worth measuring, little understanding of their relationships to one another, and little communication among researchers in different countries.

The set of acoustic properties suggested to predict annoyance is a tribute to the collective imagination of several generations of researchers. In addition to many different frequency weighting schemes, the physical properties thought to be useful in predicting annoyance have included tonality, impulsiveness, rise time, onset time, periodicity, time of day, and temporal variability. Pearsons and Bennett[5] and Schultz (see Ref. 4) have cataloged dozens of physical measures of sound that have been seriously considered as predictors of the annoyance of noise exposure.

Data dredging (misuse of regression techniques to identify and interpret random correlations among variables in large data sets) was a popular approach to inferring acoustic and other factors affecting annoyance judgments through the 1970s. Multiple regression equations so developed tended to work reasonably well for individual data sets, but poorly when applied to other data sets, so that one researcher's main effect is merely another researcher's error variance. The predictable outcome of this approach to estimating community response to noise exposure was an embarrassment of riches—not one preferred metric of noise exposure but rather an alphabet soup of measures. Botsford's[6] summary of annoyance prediction scales (which he called the "weighting" game) identifies conflicting interests in prolonging the search for the holy grail of the acoustic noise measurement procedure: the one that perfectly predicts the annoyance of a wide range of noise exposure on the sole basis of its acoustic content.

Two observations may be made about predicting the annoyance of noise exposure on the basis of measurements of properties of sound alone. The first observation is that most noise metrics correlate highly with each other, if not with self-reports of annoyance.[7] This high correlation among the various noise metrics is the basis for a common view that the choice of metric for measuring community noise is relatively unimportant. Adoption of any reasonable noise metric for regulatory purposes would have been more productive than decades of research searching for an optimal measure. For this reason, agreement by U.S. federal agencies on a set of equivalent energy noise metrics (SEL, L_{eq}, L_{dn}) in the early 1970s was an important accomplishment. This agreement was paced by the requirements of the National Environmental Policy Act in 1969 and the Noise Control Act of 1972.

The second observation is that a search for a single, purely physical metric that can serve as a universal predictor of the annoyance of noise exposure is unlikely to succeed. There are as many ways to measure noise as there are purposes for the measurement. In the absence of a theory of annoyance, the search for an optimal acoustic measure of annoyance places the cart before the horse. The fundamental problem is not, after all, measurement but rather deciding what is worth measuring. After decades of study of community reaction to noise exposure, it is now clear that it is unrealistic to expect purely physical measures to provide a complete account of the prevalence of noise-induced annoyance in communities.

Figure 1 illustrates the futility of the search for the perfect acoustic predictor of community response to noise exposure. The range and variance in this data set are so great that the shape of a dosage–response relationship intended to summarize these findings is not greatly constrained by proximity to the data points.

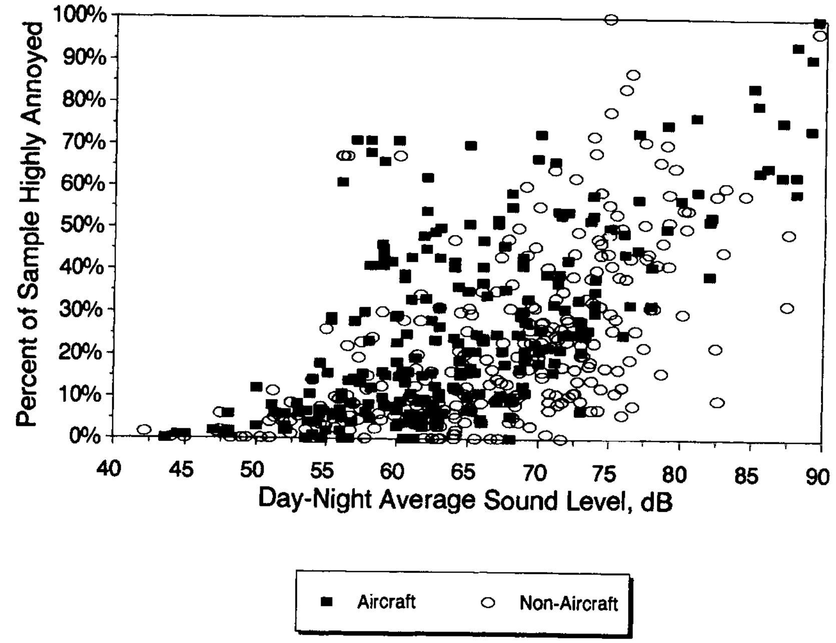


Fig. 1 Results of 545 determinations of the prevalence of noise-induced annoyance in communities exposed to transportation noise.

3 METHODS OF ASSESSING ANNOYANCE

3.1 Direct Measurement of the Prevalence of Annoyance

The least ambiguous way to estimate the prevalence of noise-induced annoyance in a community is to measure it directly by means of a social survey. Survey techniques for such purposes are well developed; Fields[8] has documented hundreds of surveys of varying sophistication on community noise effects. Considerable care must be taken in survey design, sampling, questionnaire construction, administration of interviews, and in data analysis to avoid potential errors and biases in reaching inferences on the basis of survey findings. Fidell and Green[9] provide a brief overview of the major procedural concerns in the conduct of noise effects surveys.

Less direct methods of estimating the prevalence of annoyance due to noise exposure in a community (described in the next sections) are available if the conduct of a well-designed social survey is unaffordable or otherwise impractical.

3.2 Prediction of Annoyance via an Empirical Dosage–Response Relationship

Schultz's[10] production of a well-documented, widely acknowledged quantitative dosage–response relationship via meta-analysis was a major step toward development of a standard method for prediction of transportation noise effects. Schultz fitted a third-order polynomial to a data set relating day–night average sound level (DNL) to the prevalence of a consequential degree of annoyance in communities. A simple quadratic fitting function[11] provides an updated but still purely empirical basis for estimating the prevalence of annoyance in communities:

$$\%\ \text{Highly annoyed} = 0.036L_{\text{dn}}^2 - 3.27L_{\text{dn}} + 79.14. \quad (1)$$

This relationship is simply an atheoretical fitting function that produces meaningless predictions when evaluated outside the range of approximately 45 dB $< L_{\text{dn}}$ < 85 dB. The Federal Interagency Committee on Noise prefers a logistic fit to a subset of the data reported by Fidell, Barber, and Schultz (see Ref. 11) that omits observations from several studies in which relatively low levels of noise exposure were associated with relatively high proportions of reported annoyance:

$$\%\ \text{Highly annoyed} = 100/[1 + \exp(11.13 - 0.141L_{\text{dn}})]. \quad (2)$$

This function predicts somewhat lower levels of annoyance at lower noise exposure levels than the quadratic fit: For example, a prevalence rate of only 12.3% is predicted at $L_{\text{dn}} = 65$ dB, in contrast with 18.8% for the quadratic fit to the full data set.

A CHABA Working Group on Assessment of Community Response to High-Energy Impulsive Sounds (Fidell[12]) has identified two dosage–response relationships for predicting the annoyance of impulsive sounds with CSEL values in excess of 85 dB (75 dB at night). One of these predicts the proportion of a community highly annoyed by high-energy impulsive noise exposure as

$$P = \exp(-10^{0.045(D^* - L_{\mathrm{Cdn}})}) \qquad (3)$$

where D^* is the constant 61.1 chosen to minimize the variance of the set of field observations of the annoyance of high-energy impulsive noise exposure.

An alternative method recognized in the CHABA report for predicting the prevalence of a consequential degree of annoyance due to impulsive noise exposure requires a level-dependent summation of CSEL values of individual, C-weighted impulsive noise events in annoyance units that are numerically equal to units of sound exposure for nonimpulsive sounds, measured in A-weighted units. Readers are referred to Schomer[13] for further details of this method.

Limitations of the "Equal Energy Hypothesis" as a Predictor of Annoyance Day–night level was adopted after much discussion in the 1970s by many U.S. federal agencies as a convenient general descriptor of long-term environmental noise exposure, particularly in residential settings. Although DNL was intended originally as a convenient, single-number index of complex community noise environments, its use as a predictor (and sometimes even a surrogate) of annoyance was not long delayed. This latter use of DNL has probably become its major use; after all, if noise exposure did not bother people, there would be little purpose in community noise measurement for its own sake, and little reason for measuring noise at all.

The DNL index is a time-weighted average (in effect, the average acoustic energy per second with an arbitrary nighttime weighting) variant of a family of so-called integrated energy metrics,* all of which are equally sensitive to the duration and magnitude of individual noise events and directly proportional on an energy (10 log N) basis to number of events. Reliance on any integrated energy metric to predict or explain individual and community response to aircraft noise is based on the "equal energy" hypothesis.

The equal energy hypothesis expresses the notion that the number, level, and duration of noise events are fully interchangeable determinants of annoyance as long as their product (energy summation) remains constant. Thus, whether acknowledged or not, quantification of noise exposure in units of DNL for purposes of predicting annoyance reflects a tacit theory: People are indifferent between the annoyance of small numbers of very high level noise events of short duration and the annoyance of large numbers of compensatingly lower level noise and/or longer duration noise events. The hypothesis is the underpinning of a convenient method for measuring noise exposure for purposes of predicting annoyance.

The equal energy hypothesis has provided an adequate account for data on the prevalence of annoyance due to more or less continuous (e.g., urban) noise exposure over a range of roughly 20 dB, from about L_{dn} = 55 dB to L_{dn} = 75 dB. The plausibility of the hypothesis is clearly limited in extreme cases, however. For example, no community is likely to tolerate even infrequent operation of a noise source powerful enough to damage hearing. Likewise, it is difficult to assert without substantial empirical evidence that the prevalence of long-term annoyance in a community can be predicted by the simple expedient of annualizing the exposure created by rare or infrequent noise events. Doubts about the limits of the equal energy hypothesis are often expressed as challenges to the universality of DNL as a predictor of annoyance.

Although DNL predicts the prevalence of annoyance over a 20-dB range of exposure levels with useful precision (accounting for about half of the variance in the data set), there is no evidence that the relationship between DNL and the prevalence of annoyance in a community is a causal one. Since most other reasonable metrics of noise exposure are highly correlated with DNL in airport neighborhoods, and since no other metric of noise exposure alone accounts for appreciably more variance, there are ample grounds for accepting the equal energy hypothesis in urban neighborhoods.

The utility of DNL as a predictor of the prevalence of annoyance also depends on a number of additional tacit assumptions that are most relevant to urban areas about the nature and circumstances of noise exposure. For example, reliance on equivalent energy descriptors to predict the annoyance associated with sporadic (i.e., infrequent, unexpected, and intermittent) operation of noise sources strains the hypothesis considerably. Furthermore, when noise exposure is composed of small numbers of noise intrusions of very different levels, the maximum level of a single event can control the daily integrated exposure. Thus, when dealing with truly sporadic noise exposures, it becomes essentially irrelevant whether people's reactions are based on a perfect integration of exposure, sensitivity to the maximum level, or anything in between.

---

*Other members of this family of noise measures include sound exposure level (SEL), equivalent continuous level L_{eq} and variants of DNL such as community noise equivalent level (CNEL) and C-weighted DNL. Further detail about these and other metrics in common use for characterizing environmental noise may be found in the various parts of the American National Standard "Quantities and Procedures for Description and Measurement of Environmental Sound" (ANSI S12-9).

Common sense suggests that the integration of individual noise exposures on which long-term annoyance judgments are based cannot be in perfect, loss-free proportion to the exposure created by multiple noise intrusions. If people based their annoyance judgments on a perfect accumulation of the noise exposure created by multiple events over months and years, then everyone would eventually find continued noise intrusions intolerable, and would remain fully annoyed indefinitely, even following a lengthy hiatus in exposure. Since long-term annoyance asymptotes even in heavily noise-exposed areas, the temporal integration of exposure on which annoyance judgments are based must be "leaky" in some sense. In other words, an individual's annoyance cannot ratchet indefinitely, but must at any point in time represent a balance between a rate of accumulation and a rate of dissipation.

Other Limitations of Empirical Prediction Methods None of the empirical dosage–response relationships is grounded in detailed understanding of the mechanisms by which noise exposure generates annoyance. None is fully applicable to all situations in which residential or other populations are exposed to noise. None accounts for the greater part of the variance in the social survey data.

Furthermore, an empirical dosage–response relationship does not in itself provide answers to the following sorts of questions:

- How can the prevalence of annoyance in two communities exposed to the same level of noise differ greatly?
- How accurately can the prevalence of noise-induced annoyance be predicted from knowledge of noise exposure alone?
- Decibel for decibel, is noise exposure produced by different sources equally annoying?
- What are the relative contributions of acoustic and other factors to the prevalence of annoyance in a community?
- How is the immediate annoyance created by individual events linked to the long-term annoyance of cumulative noise exposure?
- What are the time constants of arousal and decay of community annoyance?

Incomplete understanding of the origins of community annoyance can have costly consequences, including adoption of potentially ineffective means for mitigating effects of noise exposure on communities. Fidell and Silvati,[14] for example, have shown that there is no clear evidence for a reduction in the prevalence of long-term aircraft noise annoyance among homeowners whose dwellings had been treated as part of an airport noise mitigation program to increase their acoustic transmission loss. In fact, one paradoxical consequence of acoustically treating a portion of the heavily noise-exposed housing stock in the immediate vicinity of an airport could be exposure of greater numbers of untreated homes to yet higher noise exposure. This situation could arise if the numbers of flight operations increase following relief from the threat of aircraft noise litigation following relatively inexpensive acquisition of aviation easements (via home insulation projects) in neighborhoods near airports.

3.3 Theory-Based Prediction of Prevalence of Annoyance

An alternate approach to predicting the prevalence of annoyance in communities[15,16] treats annoyance not as an inexplicable concomitant of noise exposure but rather as a rational process, capable of description if not understanding in probabilistic terms. This approach permits inference of the shape of a dosage–response relationship from first principles, rather than in a purely ad hoc manner.

Fidell et al.[15] view noise exposure as a form of treatment administered to a community. The response of the community to the treatment is determined by applying a criterion for reporting annoyance to the average dose. The effective dose produced by exposure is assumed, as in the CNR system, to grow at a rate similar to the growth of loudness with level. Variability in the prevalence of annoyance in communities that is not accounted for by this compressive transform on DNL is attributed to response bias. Among the many nonacoustic factors that may affect the prevalence of noise-induced annoyance in communities are (a) various attitudes toward noise sources and their operators (approval, fear, distrust, etc.), (b) socioeconomic levels of individuals, and (c) economic dependence on operation of noise sources.

This approach follows from recognition of two separable components in self-reports of annoyance: a component directly linked to noise exposure and an entirely independent component associated with individual willingness to describe oneself as annoyed in some degree. The latter component is referred to as *response bias*. The two components are confounded in a verbal report of the form "I'm very annoyed by that aircraft flyover," since the self-report alone provides no way to distinguish the contributions of the acoustically related factors from the contributions of response bias to the expressed degree of annoyance. For lack of adequate means of distinguishing the two components, most interpretations of self-reports of annoyance tacitly assume that the contributions of response bias to annoyance are negligible.

Given that acoustic and nonacoustic determinants are confounded in individual self-reports, it follows that they are also confounded in the proportion of respondents in a neighborhood who describe themselves as highly annoyed by noise exposure. Two communities in which 20% of the residents describe themselves as highly annoyed might have quite different noise exposures. Such differences are attributed to difference in response biases among the residents. Civic action groups, political or media attention, and other factors can make the residents of the quieter community more likely to describe themselves as highly annoyed even by relatively low noise exposures.

In communities in which the prevalence of annoyance is affected primarily by noise exposure, reductions in noise exposure can be expected to lead to reductions in prevalence of annoyance. In communities in which the prevalence of annoyance is controlled by nonacoustic factors, there may be little or no reduction in annoyance associated with reductions in noise exposure.

More specifically, reactions of individuals in the community to noise exposure are assumed to be exponentially distributed with a *mean population value*, *m*. The value of *m* is assumed to be related to the day–night average sound level by

$$10 \log m = 0.3L_{\text{dn}}. \tag{4}$$

Thus, noise exposure creates a distribution of reactions within a community with a mean value that increases with the level of noise exposure. Individuals describe themselves as highly annoyed when their reactions to noise exposure exceed a fixed value of a criterion value (A) for reporting annoyance. The net effect of the nonacoustic factors on the decision-making process may be regarded as a form of response bias. The proportion of the population describing itself as highly annoyed is predicted as follows. Suppose

$$P = e^{-(A/m)}, \tag{5}$$

where P is the probability of reporting high annoyance, m is defined as in Eq. (2), and A is the criterion value for reporting annoyance. The value of A may vary from neighborhood to neighborhood for any of a number of nonacoustic reasons. For example, this criterion value may differ because the residents of one neighborhood value the commerce or convenience associated with operation of a noise source more highly than residents of another neighborhood, or because greater media or political attention has been focused on environmental problems in one neighborhood than in another, or because nonenvironmental problems are more pressing to residents of one neighborhood than of another, and so forth.

The criterion value for reporting high annoyance, D, may also be expressed in exposure-like units. If D is defined as

$$D = \frac{1}{0.3}\, 10 \log A, \tag{6}$$

then Eq. (4) may be rewritten as

$$10 \log(-\ln P) = 0.3(D - L_{\text{dn}}) \tag{7}$$

or as

$$P = \exp\,[-10^{0.03(D - L_{\text{dr}})}], \tag{8}$$

where 10 log D, or D^*, is thus a decibel-like quantity that represents the average value of the criterion for reporting annoyance in units directly comparable with L_{dn} values.

The average value of the criterion for reporting high annoyance can also depend on the source of the noise intrusion. For example, people seem to adopt a more stringent criterion for reporting annoyance due to noise from diffusely owned noise sources, such as surface street traffic, and a less stringent criterion for reporting annoyance due to aircraft operations (see Ref. 16). Thus, L_{dn} values created by aircraft and surface transportation noise can differ by 5 dB without affecting the percentage of residents who describe themselves as highly annoyed.

Equation (8) shows that if L_{dn} = 60 dB and the criterion for annoyance D^* = 70 for aircraft, then the 13.6% of the population exposed to aircraft noise will report high annoyed. If D^* = 75 for surface transportation noise, only 6% of the residents with equivalent traffic noise exposure will describe themselves as highly annoyed.

4 SPECIAL SITUATIONS

4.1 Dynamics of Community Response to Noise Exposure

No widely applicable methods are available for predicting the prevalence of noise-induced annoyance in communities in other than steady-state conditions, since little is known about the rates at which communities accommodate temporal variability in noise exposure. Although many changes in community noise exposure are gradual, step changes can occur, especially in noise produced by aircraft operations. While short-term opinions may track large exposure transients with little latency, it is clear that people distinguish between short- and long-term consequences of noise exposure. Fidell and co-workers[17] have shown that the prevalence of long-term annoyance only sluggishly reflects changes in noise exposure. At least tens of days must elapse before short-term opinions about noise exposure are transformed into long-term attitudes.

There are limits to the usefulness of viewing people as sound level meters with enormous integration times, however, particularly when exposure is intermittent and dominated by occasional high-level events. Fidell, Sneddon, and Green[18] suggest an alternative approach to modeling the annoyance of this sort of intermittent noise exposure, based on an integration of reactions to noise exposure rather than integration of noise exposure per se.

4.2 Annoyance of Low-Level Noise Intrusions

The influence of ambient noise levels on annoyance are generally neglected in assessments of the effects of high-level noise intrusions on communities. In airport neighborhoods, for example, levels of individual aircraft overflights often exceed ambient levels by 30 dB or more. In other settings, such as intrusions of en route aircraft noise in quiet suburban or rural settings, ambient noise levels can have a considerable influence on the audibility and noticeability of noise intrusions.

The audibility of a steady-state, broadband acoustic signal is determined by its bandwidth-adjusted signal-to-noise ratio, as described in many studies.[19-24] Since audibility is a matter of signal-to-noise ratio rather than signal level alone, a signal of low audibility is not necessarily a signal of low absolute level. Furthermore, the same signal can be differentially audible in different ambient noise environments.

The common unit of measurement of audibility is the scalar (dimensionless) quantity known as d'. [The audibility of an acoustic signal is given in terms of its bandwidth-adjusted signal to noise ratio: $d' = \eta S/N\ (BW)^{1/2}$. Such calculations can be conducted in one-third octave bands for most community noise applications.] An acoustic signal characterized by a d' value of 0 is physically impossible to detect, since no information is available with which to discriminate it from the background noise environment in which it occurs. Under controlled laboratory conditions, in which well-trained observers pay close attention to a structured acoustic detection task, a signal with a d' value of 2.3 will be reliably reported half of the time. A "reliable" report of the presence of a signal is one with a low false alarm rate: in the present context, a false alarm rate of 1% or less. [A "false alarm" is a report of the presence of a signal that has not in fact occurred. False alarms and correct detections co-vary in a specifiable manner for any observer (cf. Ref. 23).] Under the same highly controlled listening conditions, the presence of signals characterized by d' value of 4 is reported consistently with a negligible false alarm rate.

Under less constrained listening conditions, acoustic signals must attain slightly higher signal-to-noise ratios before observers reliably report their presence. A d' value of about 5 suffices for field observers actively listening for sound sources, while an additional 10 dB of signal-to-noise ratio ($d' = 50$) is a reasonable estimate of the audibility required to bring a signal to the attention of someone who is not actively listening for it.[25]

5 REGULATORY AND POLICY INTERPRETATIONS OF NOISE EXPOSURE EFFECTS

The pressure for standard prediction techniques and nationwide guidelines for interpreting noise exposure effects in the United States became intense following passage of the National Environmental Policy Act of 1969 and the Noise Control Act of 1972. More recent concerns in prediction of the annoyance of en route aircraft noise have arisen due to increased low-altitude, high-speed military flight activity in rural areas: U.S. Public Law 100-91 mandates assessment of impacts in nonresidential, short-term, outdoor recreational settings; development of advanced military and civil hypersonic aircraft; and unexpectedly intense adverse reactions to commercial air traffic noise in low ambient noise residential communities.

A number of organizations have developed methods and policies for community noise measurement, recommendations for land-use planning, general guidelines for noise mitigation, and various interpretations of noise effects on communities. These generic policies, recommendations, and guidelines have been developed for nationwide application and are neither intended nor guaranteed to be applied to any specific community. Guidance of this sort is fully useful only in settings unconstrained by existing land uses, such as a site selection study for a new airport.

In practice, when applied to an existing community with established land uses, the recommended planning and noise mitigation measures serve largely to render a community compatible with an airport (rather than vice versa). For example, rezoning residential land to commercial or industrial uses might in fact reduce the size of a heavily noise-impacted population. However, such rezoning might also affect a community's character, composition, quality of life, and tax base in ways that some would consider intolerable.

The best known and most well established of noise effects guidelines in the United States are those of the Federal Inter-Agency Committee on Noise (FICON).[26,27] FICON is composed of several federal agencies with interests in environmental noise exposure, including some with aviation-related interests (i.e., FAA and DoD). FICON's guidelines and recommendations for "land-use compatibility" are expressed as ranges of values of DNL. Identification of DNL values associated with "acceptable" and "unacceptable" noise exposure levels are based not on acoustic considerations per se, but rather on the expected tolerance of community effects of noise exposure on communities and on political and economic concerns. In the absence of an explicit and systematic supporting rationale for the selection of particular values of DNL as "compatible" or "incompatible" with particular land uses, this is an oddly circular approach: Both the cause (noise exposure) and the effect ("incompatibility") are defined in the same units.

FICON considers noise exposure levels lower than $L_{\mathrm{dn}} = 65$ dB to be "compatible with most residential land uses." However, FICON also states that "for populated areas, there may be appreciable numbers of persons highly annoyed by exposure below $L_{\mathrm{dn}} = 65$ dB." FICON then draws the conclusion that "thus, evaluation of the noise impact in such areas, in terms of highly annoyed, may be appropriate." In other words, if appreciable numbers of people are exposed to levels below $L_{\mathrm{dn}} = 65$ dB, it may be "appropriate" to characterize and evaluate the effects of noise exposure upon them not in acoustic terms but rather in terms of annoyance.

One implication of FICON's statement that noise exposure levels lower than $L_{\mathrm{dn}} = 65$ dB are "compatible with most residential land uses" is that it is reasonable for roughly 15% of a residential population to be highly annoyed by noise exposure [cf. Eqs. (1) and (2)]. Although it may be acceptable from some perspectives for 15% of a community to be highly annoyed by aircraft noise exposure, it may not be acceptable from other perspectives.* The U.S. Environmental Protection Agency[28] (p. 25) notes that "decisions about how much noise is too much noise for whom, for how long, and under what conditions demand consideration of economic, political, and technological matters." EPA goes on to note "the need to reconcile local economic and political realities with scientific information. People who formulate local noise abatement programs cannot escape the responsibility of making such economic and political compromises for their constituencies."

Acknowledgments

Portions of the text of this chapter paraphrase material from Fidell[29,30] Fidell and Green,[9] Fidell et al.[11,15] and Green and Fidell.[16]

REFERENCES

1. H. Fletcher, A. H. Beyer, and A. B. Duel, "Noise Measurement in City Noise," Report of the Noise Abatement Commission, Department of Health, City of New York, 1930.
2. H. Fletcher and W. A. Munson, "Loudness, Its Definition, Measurement and Calculation," *J. Acoust. Soc. Am.*, Vol. 5, 1933, pp. 82–108.
3. W. A. Rosenblith, K. N. Stevens, and the staff of Bolt Beranek and Newman, *Handbook of Acoustic Noise Control, Vol. 2, Noise and Man*, WADC TR-52-204, Wright Air Development Center, Wright-Patterson Air Force Base, Ohio, 1953.
4. T. J. Schultz, *Community Noise Rating*, Elsevier, New York, 1982.
5. K. S. Pearsons and R. Bennett, *Handbook of Noise Ratings*, NASA CR-2376, Washington, D.C., 1974.
6. J. H. Botsford, "Using Sound Levels to Gauge Human Response to Noise," *J. Sound Vib.*, Vol. 3, No. 10, 1969, pp. 16–28.
7. S. Fidell, "Community Response to Noise," in D. M. Jones and A. J. Chapman (Eds.), *Noise and Society*, Wiley, Chichester, 1984, Chap. 10.
8. J. M. Fields, *An Updated Catalog of 318 Social Surveys of Residents' Reactions to Environmental Noise (1943–1989)*, NASA TM-187553, 1991.
9. S. Fidell and D. M. Green, "Noise-Induced Annoyance of Individuals and Communities," in C. Harris (Ed.), *Handbook of Acoustical Measurements and Noise Control*, 3rd ed., McGraw-Hill, New York, 1991, Chap. 23.
10. T. J. Schultz, "Synthesis of Social Surveys on Noise Annoyance," *J. Acoust. Soc. Am.*, Vol. 64, No. 2, 1978, pp. 377–405.
11. S. Fidell, D. Barber, and T. Schultz, "Updating a Dosage-Effect Relationship for the Prevalence of Annoyance Due to General Transportation Noise," *J. Acoust. Soc. Am.*, Vol. 89, No. 1, 1991, pp. 221–233.
12. S. Fidell (Ed.), "Community Response to High-Energy Impulsive Sounds: An Assessment of the Field Since

---

*FICON explicitly acknowledges that its recommendations are merely advisory and that its policy perspectives may differ from those of local communities. In fact, FICON does not consider its guidelines as thresholds of noise impacts and encourages local government agencies to consider economic, technical, and political ramifications of noise impacts. At least one community's land-use planning criteria that are more stringent than those recommended by FICON has survived challenge in a U.S. federal appeals court.

1981," Report of CHABA Working Group 102, National Academy Press, Washington, DC, 1996.

13. P. Schomer, "New Descriptor for High-Energy Impulsive Sounds," *Noise Control Eng. J.*, Vol. 2, No. 5, 1994, pp. 179–191.
14. S. Fidell and L. Silvati, "An Assessment of the Effect of Residential Acoustic Insulation on Prevalence of Annoyance in an Airport Community," *J. Acoust. Soc. Am.*, Vol. 89, No. 1, 1991.
15. S. Fidell, T. Schultz, and D. M. Green, "A Theoretical Interpretation of the Prevalence Rate of Noise-Induced Annoyance in Residential Populations," *J. Acoust. Soc. Am.*, Vol. 84, No. 6, 1988.
16. D. M. Green and S. Fidell, "Variability in the Criterion for Reporting Annoyance in Community Noise Surveys," *J. Acoust. Soc. Am.*, Vol. 89, No. 1, 1991, pp. 234–243.
17. S. Fidell, R. Horonjeff, J. Mills, E. Baldwin, S. Teffeteller, and K. Pearsons, "Aircraft Noise Annoyance at Three Joint Air Carrier and General Aviation Airports," *J. Acoust. Soc. Am.*, Vol. 77, No. 3, 1985, pp. 1054–1068.
18. S. Fidell, M. Sneddon, and D. M. Green, "Relationship Between Short and Long Term Annoyance of Noise Exposure," USAF NSBIT TOR No. 22, July 1990.
19. S. Fidell, "Effectiveness of Audible Warning Signals for Emergency Vehicles," *Human Factors*, Vol. 20, No. 1, 1978, pp. 19–26.
20. S. Fidell, K. Pearsons, and R. Bennett, "Prediction of Aural Detectability of Noise Signals," *Human Factors*, Vol. 16, No. 4, 1974, pp. 373–383.
21. S. Fidell, S. Teffeteller, R. Horonjeff, and D. M. Green, "Predicting Annoyance from Detectability of Low Level Sounds," *J. Acoust. Soc. Am.*, Vol. 66, No. 5, 1979, pp. 1427–1434.
22. S. Fidell and S. Teffeteller, "Scaling the Annoyance of Intrusive Sounds," *J. Sound Vib.*, Vol. 78, No. 2, 1981, pp. 291–298.
23. D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics*, Wiley, New York, 1966.
24. L. Hutchings and S. Fidell, "Documentation for Version 7 of Tank-Automotive Command's Acoustic Detection Range Prediction Model (ADRPM-7)," Warren, MI, 1989.
25. S. Fidell, K. Pearsons, and M. Sneddon, "Evaluation of the Effectiveness of SFAR 50-2 in Restoring Natural Quiet to Grand Canyon National Park," BBN Report 7197, February 1992.
26. American National Standard S12.40-1990, "Sound Level Descriptors for Determination of Compatible Land Use," Standards Secretariat, Acoustical Society of America, New York, August 1990.
27. Federal Interagency Committee on Noise (FICON), "Final Report: Airport Noise Assessment Methodologies and Metrics," Washington, D.C., 1992.
28. United States Environmental Protection Agency, "Protective Noise Levels," Condensed Version of EPA Levels Document, EPA 550/9-79-100, November 1978.
29. S. Fidell, "Community Response to Noise," in C. Harris (Ed.), *Handbook of Noise Control*, 2nd ed., McGraw-Hill, New York, 1976, Chap. 36.
30. S. Fidell, "An Historical Perspective on Predicting the Annoyance of Noise Exposure," presented at NOISE-CON 90, Austin, Texas, October 15–16, 1990.